SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

TOP SCIENCE STORIES OF 2018

Hawking's legacy

The new history of dinosaurs

Can we spot fake videos?

Kilauea blows its top

The Arctic meltdown heats up

Liquid water found on Mars

WINTER 2018/2019

© 2018 Scientific American

Expertise. Insights. Illumination.

Discover world-changing science. Get 12 issues of *Scientific American* in print and explore our digital archive back to 1845, including articles by more than 150 Nobel Prize winners.

sciam.com/print&archive

Download on the App Store | GET IT ON Google Play

SPECIAL ISSUE : SEX AND GENDER
GENDER MYTHS DEBUNKED
UNHEALTH PRACTICE
SPEAKING IN WHISTLES
WHAT'S WRONG WITH COSMIC INFLATION?
SCIENTIFIC AMERICAN
WORKOUTS AND WEIGHT LOSS
Learn the surprising evolutionary reason why exercise alon won't shed pounds—and what to do about it
PLUS
PREVENTING THREATS TO CROPS
Airborne microbes' critical role PAGE 40
WHAT IS A KILOGRAM?
Science seeks new ways to ensure accuracy PAGE 46
MEDICAL SCANS
A shortage of rare atoms imperils disease imaging PAGE 68
SCIENTIFIC AMERICAN
AUTOMOBILE NUME
SCIENTIFIC AMERICAN
T. REX reexamined
SCIENTIFIC AMERICAN
September 1999

Scientific American is a registered trademark of Springer Nature America, Inc. Google Play and the Google Play logo are trademarks of Google LLC. Apple, the Apple logo, iPhone, and iPad are trademarks of Apple Inc., registered in the U.S. and other countries and regions. App Store is a service mark of Apple Inc.

SCIENTIFIC AMERICAN™
ESTABLISHED 1845

Top Science Stories of 2018 is published by the staff of Scientific American, with project management by:

Editor in Chief and Senior Vice President: **Mariette DiChristina**

Managing Editor: **Curtis Brainard**

Collections Editor: **Andrea Gawrylewski**

Creative Director: **Michael Mrak**

Issue Designer: **Lawrence R. Gendron**

Photography Editor: **Monica Bradley**

Assistant Photo Editor: **Liz Tormes**

Copy Director: **Maria-Christina Keller**

Senior Copy Editor: **Daniel C. Schlenoff**

Copy Editor: **Aaron Shattuck**

Managing Production Editor: **Richard Hunt**

Prepress and Quality Manager: **Silvia De Santis**

Editorial Administrator: **Ericka Skirpan**

Senior Secretary: **Maya Harty**

Advertising Production Controller: **Carl Cherebin**

Production Controller: **Madelyn Keyes-Milch**

President: **Dean Sanderson**

Executive Vice President: **Michael Florek**

Vice President, Commercial: **Andrew Douglas**

Publisher and Vice President: **Jeremy A. Abbate**

Associate Vice President, Business Development: **Diane McGarvey**

Head, Marketing and Product Management: **Richard Zinken**

Marketing Director, Institutional Partnership and Customer Development: **Jessica Cole**

Online Marketing Product Manager: **Zoya Lysak**

Digital Marketing Manager: **Marie Cummings**

E-Mail Marketing Manager: **Chris Monello**

Marketing and Customer Service Coordinator: **Christine Kaelin**

Head, Communications, USA: **Rachel Scheer**

Director, Integrated Media: **Jay Berfas**

Director, Integrated Media: **Matt Bondlow**

Custom Publishing Editor: **Lisa Pallatroni**

Rights and Permissions Manager: **Felicia Ruocco**

Senior Administrator, Executive Services: **May Jung**

COVER IMAGES (*counterclockwise from top left*): MICHELLY RALL *Getty Images (Hawking)*; TAYLOR CALLERY *(digital eye)*; GETTY IMAGES *(ice cap)*; NASA/JPL/MALIN SPACE SCIENCE SYSTEMS *(Mars)*; MARIO TAMA *Getty Images (Kilauea)*; RICARDO N. MARTINEZ *Division of Vertebrate Paleontology Institute and Museum of Natural Sciences, National University of San Juan (Saurosuchus)*; THIS PAGE: BRUNO VINCENT *Getty Images*

Stephen Hawking appears at the release of the *Bulletin of the Atomic Scientists* on January 17, 2007, in London.

FROM THE EDITOR

The Year in Science

In March physicist Stephen Hawking passed away after living for more than 50 years with amyotrophic lateral sclerosis, a devastating neurodegenerative disease that confined him to a wheelchair for most of his life. Revered for his description in the 1970s of the radiation emanating from black holes, he became a cultural icon, authoring best-selling books and appearing in documentaries and sitcoms as recently as March. His death spurred a global outpouring of tributes from fellow scientists, world leaders and celebrities (*page 52*). Perhaps his popularity resulted from his resilient pursuit of knowledge—a reminder of the power of science to inspire and uplift.

Several of this year's discoveries in cosmology and physics invoke that Hawking-esque wonder. In May radar observations revealed what seems to be a liquid lake underneath the surface of Mars (*page 38*). Since the detection of gravitational waves three years ago, the results have been rolling in, now combining with light measurements that help to quickly triangulate cosmic detections in a novel type of astronomy (*page 54*).

Discoveries in laboratories are no less impressive. Recent insights that the brain works mechanically rather than electrically (*page 22*) and in close cooperation with the immune system (*page 4*) could transform neuroscience; by examining wear patterns on ancient hominid teeth, paleontologists have also revealed the true human diet (hint: it's good news for foodies, bad news for fad diets) (*page 30*).

The ever changing earth kept scientists on their toes in 2018. Rising seawaters are forcing the relocation of many coastal communities and causing a societal rethink over where we build (*page 60*); a dazzling eruption of Hawaiian volcano Kilauea is yielding data that will (hopefully) tip off geologists to the next big blow, wherever it might occur (*page 80*).

Science took center stage in many societal hot topics. Innovative digital models are helping epidemiologists get a handle on the opioid crisis and ways we might combat it (*page 106*). Economist and Nobel Prize winner Joseph E. Stiglitz has some recommendations for how to rebalance a growingly inequitable society (*page 88*). Tech researchers are devising better safeguards against the next fake news culprits: videos altered using AI (*page 100*).

Hawking once wrote: "Scientists have become the bearers of the torch of discovery in our quest for knowledge." Never has it felt more true that wherever there is an unresolved question, desperate need or global phenomenon, you will find science.

Andrea Gawrylewski
Collections Editor
editors@sciam.com

© 2018 Scientific American

SPECIAL EDITION

SCIENTIFIC AMERICAN

Volume 27, Number 5, Winter 2018/2019

4 INSIGHTS INTO HEALTH AND THE MIND

38 FRONTIERS OF SPACE AND PHYSICS

4 The Seventh Sense
Long thought to be divorced from the brain, the immune system turns out to be intimately involved in its functioning. *By Jonathan Kipnis*

12 American Epidemic
Resurgent outbreaks of infectious diseases are sickening thousands, and the causes are societal. *By Melinda Wenner Moyer*

Also: **Vaccine Inequality** *By Lee Riley*

22 The Brain, Reimagined
Physicists who have revived experiments from 50 years ago say nerve cells communicate with mechanical pulses, not electric ones. *By Douglas Fox*

30 The Real Paleo Diet
Microscopic wear patterns on fossil teeth reveal what our ancestors ate—and provide insights into how climate change shaped human evolution. *By Peter S. Ungar*

38 Watery Red Planet?
Radar observations have revealed what appears to be a buried lake on Mars, the first ever stable reservoir of liquid water found on the planet. *By Lee Billings*

42 The Unsolvable Problem
Three mathematicians have discovered that a problem of central importance in physics is impossible to solve. *By Toby S. Cubitt, David Pérez-García and Michael Wolf*

52 Life of a Legend
Science mourns the death of physicist and science icon Stephen Hawking. *By Davide Castelvecchi*

54 Messengers from the Sky
Astronomers' newfound ability to see the same cosmic events in light, particles and gravitational waves—a synthesis called multimessenger astronomy—gives them a fuller picture of some of the universe's most mysterious phenomena. *By Ann Finkbeiner*

© 2018 Scientific American

60 NEW VIEWS OF LIFE ON EARTH

60 Underwater
Coastal communities struggling to adapt to rising seas are beginning to do what was once unthinkable: retreat. *By Jen Schwartz*

72 The Unlikely Triumph of Dinosaurs
New fossils and analyses topple the long-standing explanation of how dinosaurs came to rule the earth. *By Stephen Brusatte*

80 Canary in the Crater
Scientists scramble to analyze data from Kilauea, hoping to predict similar volcanic events elsewhere. *By Sara Reardon*

82 Meltdown
The Arctic climate is shattering record after record, altering weather worldwide. *By Jennifer A. Francis*

88 WHERE SCIENCE MEETS SOCIETY

88 A Rigged Economy
The American political system fosters one of the most inequitable societies among developed nations. What can we do about it? *By Joseph E. Stiglitz*

94 The World Bitcoin Created
The first big digital currency gave us a glimpse of a new economic order—one that raises more questions than answers. *By John Pavlus*

100 Clicks, Lies and Videotape
Artificial intelligence is making it possible for anyone to manipulate audio and video. The biggest threat is that we stop trusting anything at all. *By Brooke Borel*

106 Model Citizens
How digital drug dealers and virtual users are providing cludes to help stop the U.S. opioid crisis. *By Sara Reardon*

DEPARTMENTS

1 FROM THE EDITOR: The Year in Science

112 END NOTE: A Good Life for All

Articles in this special issue are updated or adapted from previous issues of *Scientific American* and from ScientificAmerican.com.
Copyright © 2018 Scientific American, a division of Springer Nature America, Inc. All rights reserved. Scientific American Special (ISSN 1936-1513), Volume 27, Number 5, Winter 2018/2019, published by Scientific American, a division of Springer Nature America, Inc., 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 1 New York Plaza, Suite 4500, New York, NY 10004-1562. **Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.**

© 2018 Scientific American

NEUROSCIENCE

THE SEVENTH SENSE

Long thought to be divorced from the brain, the immune system turns out to be intimately involved in its functioning

By Jonathan Kipnis

Illustration by Mark Ross Studios

© 2018 Scientific American

© 2018 Scientific American

FOR DECADES ANATOMY TEXTBOOKS TAUGHT THAT THE TWO MOST complicated systems in the body—the brain and the immune system—existed in almost complete isolation from each other. By all accounts, the brain focused on the business of operating the body, and the immune system focused on defending it. In healthy individuals, the twain never met. Only in certain cases of disease or trauma did cells from the immune system enter the brain, and when they did so, it was to attack.

But in recent years a rush of new findings has revolutionized scientists' understanding of the two systems. Mounting evidence indicates that the brain and the immune system interact routinely, both in sickness and in health. The immune system can help support an injured brain, for example. It also plays a role in helping the brain to cope with stress and aids such essential brain functions as learning and social behavior. What is more, the immune system might qualify as a kind of surveillance organ that detects microorganisms in and around the body and informs the brain about them, much as our eyes relay visual information and our ears transmit auditory signals. In other words, the brain and immune system do not just cross paths more often than previously thought—they are thoroughly entwined.

IN BRIEF

Conventional wisdom long held that the brain and the immune system do not interact in healthy individuals. **In recent years,** however, researchers have amassed ample evidence that the two systems are closely connected. **Scientists still have** much to learn about this new field of neuroimmunology, but its findings could lead to new insights into—and therapies for—a number of neurological diseases.

Researchers are still in the early stages of studying this burgeoning new field of neuroimmunology. But already it is becoming clear that the brain's response to immunological information and how that information controls and affects brain circuitry could be the key to understanding many neurological diseases—from autism to Alzheimer's—and developing new therapies for them. Efforts to treat such disorders have typically met with disappointing results because most drugs cannot easily penetrate the brain. The findings from neuroimmunology raise the tantalizing possibility that targeting the immune system might be a more effective tactic.

RECEIVED WISDOM

TO UNDERSTAND THE SIGNIFICANCE of these discoveries, it helps to know a bit about how the brain and immune system are structured and how they work. The brain is our supercomputer and master regulator. Working with the spinal cord and several cranial nerves, which together constitute the central nervous system (CNS), it controls all the body's functions. Given the vast scope of the brain's responsibility, it is perhaps no surprise that the organ is incredibly intricate. Its basic functional units are neurons, which occupy roughly half of the brain. The human brain contains an estimated 100 billion neurons interlinked by approximately 100 trillion connections called synapses. The neurons, along with various types of nonneuronal cells called glia, make up the brain's parenchyma, the functional tissue responsible for processing information. Other key players include stromal cells, which physically support the parenchymal tissues, and endothelial cells, which compose the blood vessels that supply the brain and form the blood-brain barrier, which limits the passage of substances from other parts of the body into the brain.

For its part, the immune system has two major components, innate immunity and adaptive immunity. Innate immunity is the more primitive element, having evolved about a billion years ago in the first cells to detect and dispatch enemy forces quickly but without much precision. It is the body's first line of defense against pathogens, consisting of physical and chemical barriers to them, as well as cells that kill them. Innate immunity initiates the inflammatory response, in which white blood cells swarm the site of infection and churn out proteins that induce heat and swelling to confine and destroy pathogens. Adaptive immunity, which evolved after the innate component, consists mainly of cells called T lymphocytes and B lymphocytes, which can recognize a specific pathogen and mount a corre-

© 2018 Scientific American

spondingly targeted attack against it. In a perfect world, all adaptive immune cells would take aim only at external pathogens and would not touch the body's own proteins or cells. But in about 1 percent of the population, adaptive immunity loses control and attacks cells in the individual's own tissues, causing autoimmune diseases such as multiple sclerosis, arthritis and certain forms of diabetes, among many others. Still, the system has an impressive success rate, targeting foreign invaders exclusively in some 99 percent of individuals.

Researchers long thought that the immune system worked by simply distinguishing an organism's own constituents from nonself ones. But eventually more complex theories began to emerge. In the 1990s Polly Matzinger of the National Institute of Allergy and Infectious Diseases proposed that the immune system recognizes not only foreign invaders but also damage to tissues. This notion gained support from the subsequent identification of molecules that are released by injured, infected or otherwise damaged tissues. These molecules attract the attention of the immune cells, triggering a cascade of events that lead to activation of the immune system, recruitment of immune cells to the site of injury, and elimination (or at least an attempt at elimination) of the alarm-causing invader or injury. In addition, experiments have found that suppression of adaptive immunity accelerates the development and growth of tumors and slows down the healing process in damaged tissues. Such findings show that the immune system—once considered to be laser-focused on protecting the body from foreign invaders—actually has a far greater purview: regulating the body's tissues to help them to maintain equilibrium in the face of all manner of insults, whether from without or within.

But until recently, scientists were quite sure that this purview did not extend to the brain. As early as the 1920s, researchers observed that although the healthy brain harbors immune cells native to the CNS called microglia, immune cells from elsewhere in the body (so-called peripheral immune cells) are not usually found there. The blood-brain barrier keeps them out. In the 1940s biologist Peter Medawar, who won a Nobel Prize for his research, showed that the body is slower to reject foreign tissue grafted onto the brain than grafts placed elsewhere in the body. The brain was "immune privileged," Medawar argued, impervious to the immune system. Peripheral immune cells do appear in the parenchyma and spinal cord of patients with brain infections or injuries, however. And mouse studies demonstrate that these cells cause the debilitating paralysis associated with the disease. Based on such findings, scientists suggested that the brain and immune system have nothing to do with one another except in cases of pathologies that allow immune cells to enter the CNS and wage war on neurons.

(Exactly how the immune cells breach the blood-brain barrier in such instances is uncertain. But it may be that the barrier gets activated during brain diseases in ways that allow immune cells to cross over. In a seminal study published in 1992, Lawrence Steinman of Stanford University and his colleagues found that in mice with a condition similar to multiple sclerosis, peripheral immune cells make a protein called $\alpha 4\beta 1$ integrin that allows them to penetrate the barrier. A drug that inhibits the interaction between the integrin and the endothelial cells, Tysabri, is one of the most potent treatments for multiple sclerosis patients.)

The theory that the brain and immune system lead separate lives prevailed for decades, but it was not without skeptics. Some wondered why, if the immune system is the body's main fighting force against pathogens, the brain would give up ready access to such a system of defense. Supporters of the theory responded that the blood-brain barrier prevents the entry of most pathogens into the brain, so the brain has no need to accommodate the immune system, especially if it could cause problems by being there—doing battle with neurons, for instance. The skeptics pointed out that several viruses, as well as some bacteria and parasites, can access the brain. And far from ignoring these transgressions, the immune system responds to them, rushing to the brain to manage the invading agent. Perhaps the scarcity of pathogens in the brain is not because the blood-brain barrier is so effective at filtering them out but because the immune system is so efficient at fighting them. Indeed, studies have shown that immunosuppressed patients suffer complications that often affect the CNS.

Mice lacking adaptive immunity showed not only impaired spatial learning behavior but also compromised social behavior.

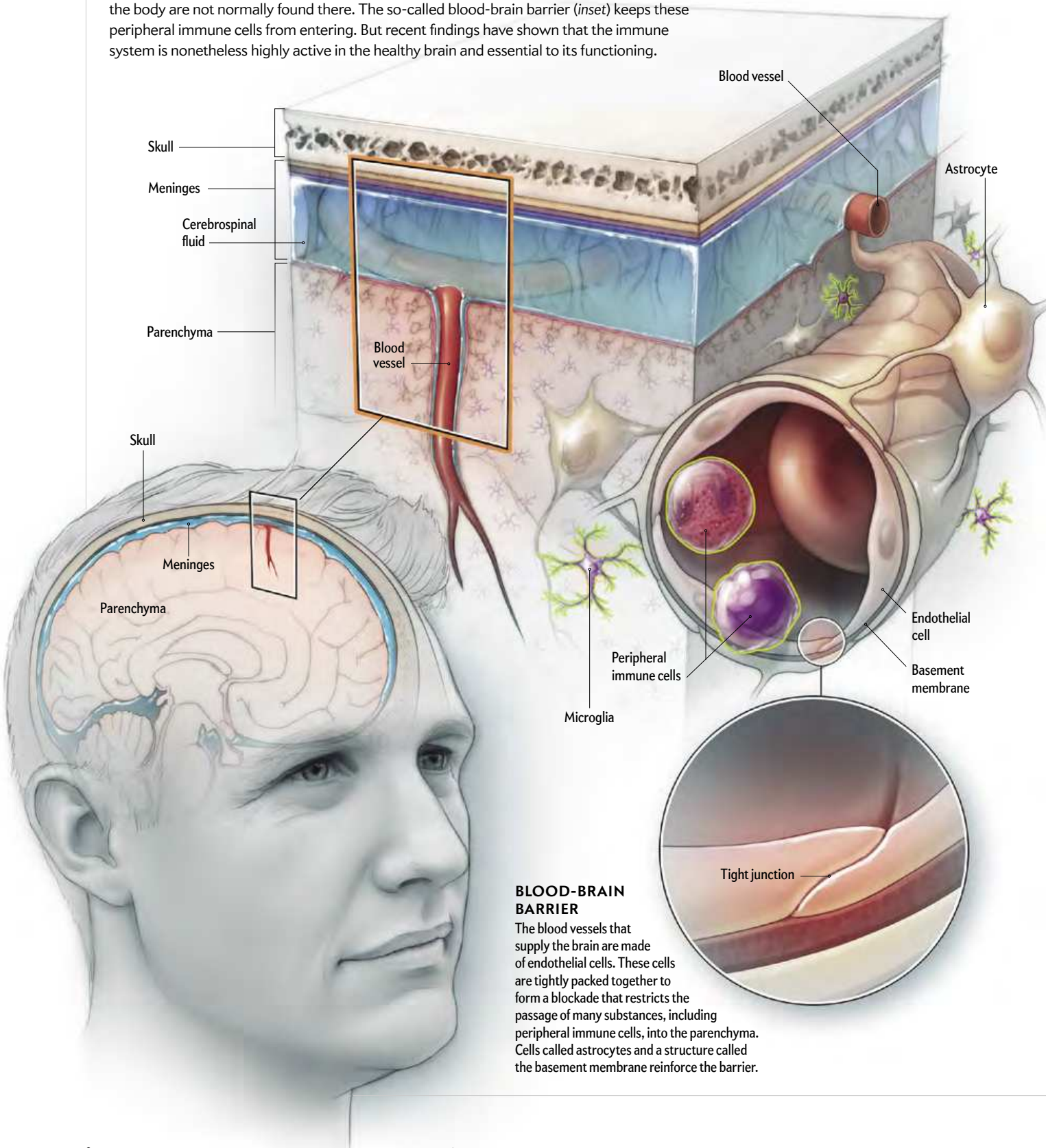
REWRITING THE TEXTBOOKS

EVENTUALLY SUCH ARGUMENTS and a growing appreciation of the immune system's role in supporting damaged bodily tissues prompted researchers to reexamine its role in the CNS. When they took a closer look at the CNS in rats and mice with spinal cord injuries, they found it overrun with infiltrating immune cells. In experiments carried out in the late 1990s, Michal Schwartz of the Weizmann Institute of Science in Rehovot, Israel, showed that eliminating immune cells after injury to the CNS worsens neuron loss and brain function, whereas boosting the immune response improves neuron survival. More recently, studies led by Stanley Appel of Houston Methodist Hospital and Mathew Blurton-Jones of the University of California, Irvine, have found that amyotrophic lateral sclerosis

© 2018 Scientific American

The Brain-Immune Connection

The healthy brain was long thought to be off-limits to the immune system. Although the brain harbors native immune cells known as microglia, immune cells that originate elsewhere in the body are not normally found there. The so-called blood-brain barrier (*inset*) keeps these peripheral immune cells from entering. But recent findings have shown that the immune system is nonetheless highly active in the healthy brain and essential to its functioning.

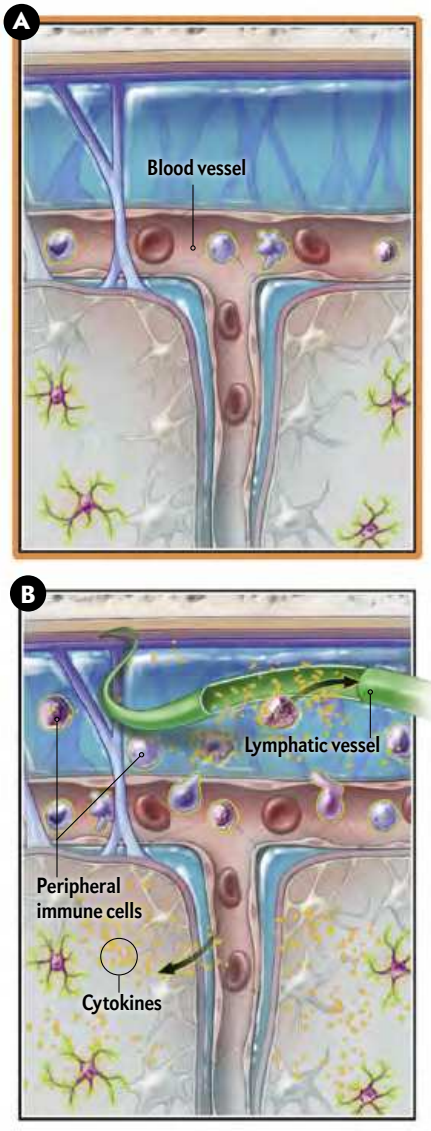


BLOOD-BRAIN BARRIER
The blood vessels that supply the brain are made of endothelial cells. These cells are tightly packed together to form a blockade that restricts the passage of many substances, including peripheral immune cells, into the parenchyma. Cells called astrocytes and a structure called the basement membrane reinforce the barrier.

© 2018 Scientific American

CIRCUMVENTING THE BARRIER

Until recently, researchers thought that the membranes surrounding the parenchyma, called the meninges, functioned mainly to carry the cerebrospinal fluid that buoys the brain **A**. New findings show there is more to the story **B**. The meninges turn out to contain lymphatic vessels that remove toxins and other waste from the parenchyma and can relay information about brain infections to the immune system. The meninges also house an array of peripheral immune cells that can communicate with the brain by means of proteins they manufacture called cytokines. Cerebrospinal fluid from the meninges enters the parenchyma through spaces surrounding the blood vessels supplying the brain and can thus carry cytokines from the peripheral immune cells deep into the brain to influence neuron behavior.



and Alzheimer's disease develop more severely and rapidly in mice engineered to lack adaptive immunity than in normal mice. Restoring adaptive immunity slows the progression of such diseases. These results indicate that immune cells help neurons rather than only hurting them, as was previously supposed.

At first glance, the immune system's intervention to protect the injured CNS does not make sense. When the CNS sustains trauma, the immune system mounts an inflammatory response, releasing toxic substances to eliminate pathogens and, in some cases, to remove damaged cells, which thereby restores equilibrium. The inflammatory response is a blunt instrument, however, taking out some of the good guys along with the bad. In other tissues, such collateral damage is tolerable because the tissues regenerate readily. But CNS tissue is limited in its ability to grow back, which means that damage from the immune response is typically permanent. Given the potential for immune activity to wreak havoc in the brain, the costs of intervention could often outweigh the benefits. But maybe the immune response observed after CNS injury is simply an extension of the immune response that aids brain function under normal conditions.

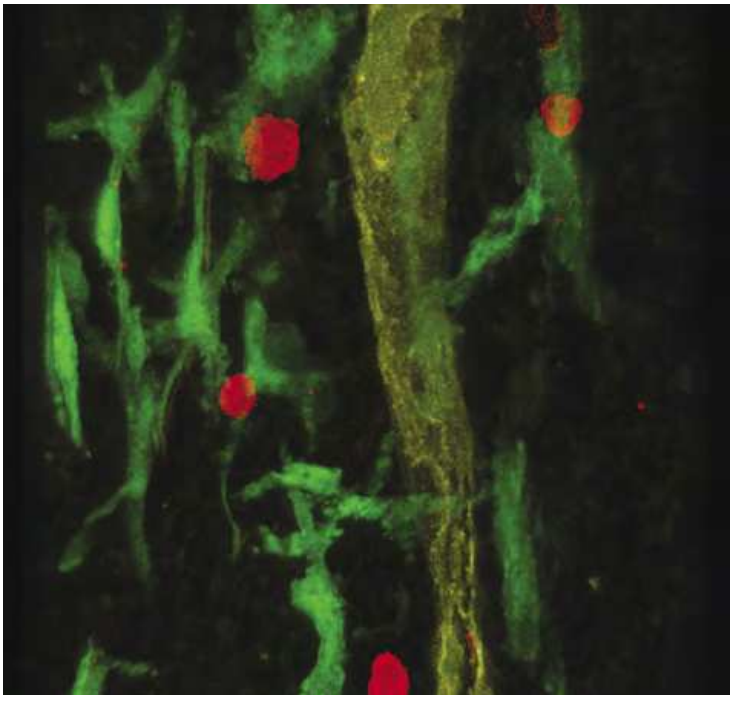
Recent studies support this notion. My collaboration with Hagit Cohen of Ben-Gurion University of the Negev in Israel and Schwartz revealed that mice that experience stressful stimuli, such as exposure to the smell of their natural predators, develop an immediate stress response—in this case, hiding in a maze rather than exploring it. In 90 percent of cases, the stress response disappears within hours or days. But for the other 10 percent, the response persists for days to weeks. Mice in the latter group can thus serve as an animal model for post-traumatic stress disorder (PTSD). Interestingly, when mice lacking adaptive immunity are compared with mice that have a normal immune system, the incidence of PTSD is increased severalfold. These results provided the first indication that the immune system supports the brain not only during infections and injuries but also during psychological stress. Moreover, some evidence links the immune system to PTSD in humans.

Though not as nerve-racking as exposure to a predator, tasks that require learning are also stressful. Think of preparing for an exam or even cooking a new recipe. Could an inability to deal with stress hinder the learning process itself? To test this hypothesis, my colleagues and I compared the performance of mice lacking adaptive immunity with that of a control group in various behavioral tests. We found that mice without adaptive immunity, unlike the controls, performed poorly in tasks requiring spatial learning and memory, such as figuring out the location of a platform hidden in a large pool of water. We have since shown that the mice lacking adaptive immunity exhibit not only impaired spatial learning behavior but also compromised social behavior, preferring to spend their time with an inanimate object rather than another mouse.

As evidence that the immune system plays important roles in different brain functions has accumulated, new unknowns have emerged. How the immune system exerts its influence in the CNS is one. After all, apart from microglia, no immune cells are present within the parenchyma of healthy individuals. Clues have come from proteins called cytokines, which are made by immune cells and influence the behavior of other cells. Cytokines released by peripheral immune cells can affect the brain. They presumably gain entrance through brain areas that lack the regular blood-brain barrier and could directly impact the brain through the vagus nerve, which runs from the brain to the abdomen. The available

Illustration by David Cheney

© 2018 Scientific American



STAINED TISSUE highlights T cells (*red*) and macrophages (*green*), as well as a lymphatic vessel (*yellow*), in the meninges.

ANTOINE LOUVEAU *University of Virginia School of Medicine*

evidence suggests that the immune cells within the meninges—the membranes that surround the brain—are also the source of the cytokines that may affect brain function. How these immune cells enter the meninges, how they circulate there and how they produce their cytokines are currently subjects of intensive research.

Recently my colleagues and I made an intriguing discovery that bears on these questions: It has to do with how the body gets rid of toxins and waste. The tissues in the body contain two types of vessels. Just as a house has two types of pipes that serve it, one for water and the other for sewage, our tissues have the blood vessels that carry oxygen and nutrients to them and the lymphatic vessels that remove toxins and other waste materials that the tissues produce. The lymphatic vessels also ferry antigens—substances capable of inducing an immune response—from the tissues into tissue-draining lymph nodes, where they are presented to immune cells to be inspected for information on the draining tissue. On detecting a problem, such as injury or infection in the tissue, the immune cells activate and migrate to the affected tissue to try to resolve the problem.

Because of the enduring belief that the healthy brain is disconnected from the immune system and because the parenchyma does not contain lymphatic vessels, scientists long assumed that neither the brain nor the rest of the CNS is serviced by the lymphatic network. Yet this assumption presented a conundrum: Why would the brain not report to the immune system about potential problems that might be affecting it and that the immune system might help solve? And how does the immune system nonetheless receive information on brain infections? Furthermore, studies have found that brain injuries provoke a strong immune response in lymph nodes located outside the brain. How is that possible?

Fascinated by the immune activity in the meninges and its effects on brain function, my colleagues and I decided to take a closer look at those membranes. In doing so, we made a serendipitous discovery: it turns out they house lymphatic vessels. Several other research groups have since made similar findings in fish, mice, rats, nonhuman primates and humans. The results confirm earlier proposals for a link between the brain and lymph system that were made some 200 years ago but largely dismissed. These vessels represent a bona fide lymphatic network that drains the CNS, a missing link that can relay information about brain infections and injuries to the immune system.

The presence of both lymph vessels and immune cells in the meninges means researchers need to rethink the exact function of these membranes. The traditional explanation holds that they simply carry the cerebrospinal fluid, which buoys the brain. But considering how densely packed the brain's constituent cells are and how sensitive its neurons are when they fire their electrical signals, perhaps moving all of the brain's immune activity to its meningeal borders was evolution's solution to the problem of allowing the immune system to serve the entire CNS without interfering with neuron function.

The discovery of the brain's lymphatic vessels revealed how the immune system receives information about tissue damage in the CNS. For insights into how the meningeal immune cells actually communicate with the parenchyma and affect it from afar, however, we have to turn to another branch of the brain's waste-removal system. In addition to the lymphatic network that we discovered, the CNS also has a network of channels in the parenchyma through which the cerebrospinal fluid gets access to the brain. Maiken Nedergaard of the University of Rochester has dubbed this network the glymphatic system. The fluid enters the parenchyma through spaces surrounding the arteries that pipe into the brain from the meninges and washes through the tissues until it is recollected in the spaces surrounding the veins and then returned to the pool of cerebrospinal fluid in the meninges. This flow of fluid presumably carries immune molecules such as cytokines from the meninges into the parenchyma, where they can exert their influence.

Studies of cytokines have illuminated how they modulate behavior. For example, Robert Dantzer, now at the University of Texas MD Anderson Cancer Center, and Keith Kelley of the University of Illinois at Urbana-Champaign have determined that interleukin-1 beta initiates sickness behavior, the name given to the constellation of behaviors people typically exhibit when ill, such as sleeping excessively, eating less and withdrawing from social contact. And my own team has recently shown that interferon gamma, a cytokine produced by

© 2018 Scientific American

meningeal T cells, interacts with neurons in the brain's prefrontal cortex, which, among its other functions, is involved in social behavior. Surprisingly, this cytokine does not exert its influence via the brain's resident immune cells (the microglia) but rather those neurons that control the circuits associated with social behavior. In fact, the cytokines are essential for proper functioning of these circuits: in the absence of T cells or their interferon gamma, these neurons fail to regulate the circuits correctly, and circuit hyperactivity ensues—a disturbance linked to social deficits. Thus, a cytokine produced by immune cells in the meninges can change the activity of neurons, thereby altering the function of the circuit and changing the underlying behavior.

Interferon gamma is not the only immune molecule that affects brain function. Mario de Bono of the MRC Laboratory of Molecular Biology in England and his colleagues have shown that another cytokine, IL-17, activates sensory neurons in the roundworm *Caenorhabditis elegans* and changes the creature's oxygen-sensing behavior. And recent work in mice by Gloria Choi of the Massachusetts Institute of Technology and her collaborators has demonstrated that IL-17 can interact with neurons in the brain's cortex and alter behaviors related to autism spectrum disorder.

ANOTHER SENSE ORGAN?

ONE MIGHT WONDER why an organ as powerful as the brain needs to be controlled or supported by the immune system to function properly. I have developed a hypothesis for why the two systems are so closely linked. We have five established senses—smell, touch, taste, sight and hearing. The sense of position and movement, or proprioception, is often referred to as the sixth sense. These senses report to the brain about our external and internal environments, providing a basis on which the brain can compute the activity needed for self-preservation. Microorganisms abound in these environments, and the ability to sense them—and defend against them when needed—is central to survival. Our immune system excels at exactly that, with innate immunity's ability to generally recognize patterns and types of invaders and adaptive immunity's talent for recognizing specific invaders. I propose that the defining role of the immune system is to detect microorganisms and inform the brain about them. If, as I suspect, the immune response is hardwired into the brain, that would make it a seventh sense.

There are ways to test this hypothesis. Because the brain's circuits are all interconnected, interference with one circuit tends affect others as well. For instance, food tastes different when our sense of smell is impaired. Evidence that interference with immune input disturbs other circuits would support the idea that the immune response is a hardwired seventh sense. One possible example comes from sickness behavior. Perhaps an overwhelming input of signals from the seventh sense informing the brain of pathogenic infection spills over and disrupts the circuits that modulate sleepiness, hunger, and so on during illness, leading to this characteristic set of behavioral changes that develop in affected individuals. Alternatively, the microorganism information relayed to the brain by the immune sensory system may prompt the brain to initiate sickness behavior as a means of protecting the sick individual by minimizing exposure to other pathogens and conserving energy.

Our knowledge of the relation between the brain and the immune system is still in its infancy. We should not be surprised if new discoveries in this field over the next 10 or 20 years reveal the two systems in a completely different light. I hope, though, that the fundamental understanding we possess today will be enriched by the results of such research rather than overturned altogether. One research priority will be mapping how the immune components and neural circuits connect, interact and interdepend in health and disease. Knowing those relations will allow investigators to target immune signaling in their treatment of neurological and mental disorders. The immune system is an easier drug target than the CNS, and it is plausible that one day repair of the immune system through gene therapy or even the replacement of a flawed immune system via bone marrow transplantation will be a viable means of treating brain disorders. Given the myriad immune alterations in brain disorders, research on neuroimmune interactions will probably continue for decades to come and gradually reveal to us even deeper mysteries of the brain. SA

If the immune response is hardwired into the brain, that would make it a seventh sense.

Jonathan Kipnis is a professor and chair of neuroscience and director of the Center for Brain Immunology and Glia at the University of Virginia School of Medicine. His research focuses on interactions between the nervous and immune systems.

MORE TO EXPLORE

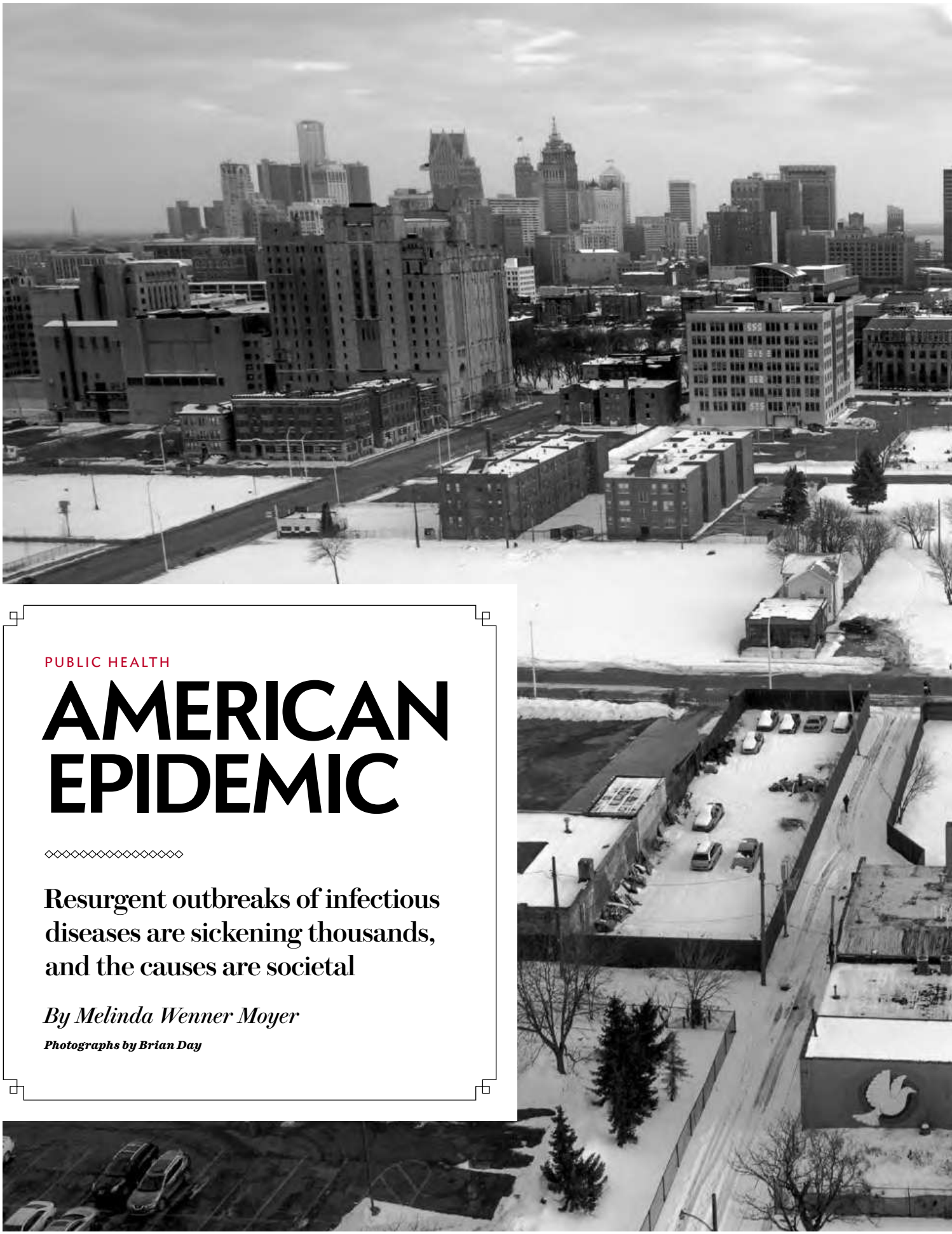
Structural and Functional Features of Central Nervous System Lymphatic Vessels. Antoine Louveau et al. in *Nature,* Vol. 523, pages 337–341; July 16, 2015.
Multifaceted Interactions between Adaptive Immunity and the Central Nervous System. Jonathan Kipnis in *Science,* Vol. 353, pages 766–771; August 19, 2016.
Immune System: The "Seventh Sense." Jonathan Kipnis in *Journal of Experimental Medicine,* Vol. 215, No. 2, pages 397–398; February 2018.

FROM OUR ARCHIVES

Brain Drain. Maiken Nedergaard and Steven A. Goldman; March 2016.

scientificamerican.com/magazine/sa

© 2018 Scientific American



PUBLIC HEALTH

AMERICAN EPIDEMIC

Resurgent outbreaks of infectious diseases are sickening thousands, and the causes are societal

By Melinda Wenner Moyer

Photographs by Brian Day

© 2018 Scientific American

MIDTOWN DETROIT'S Cass Corridor neighborhood has new construction, as well as a population of sick and homeless people.

© 2018 Scientific American

DEAN CARPENTER ZIGZAGS HIS WAY THROUGH A ROW OF MEN seated in hard plastic chairs at Detroit's Tumaini Center, a crisis support organization for the chronically homeless in Michigan's biggest city. The center has no beds, so some men have been living in those chairs for weeks, even years, while case workers try to secure them housing. Carpenter, the center's nurse practitioner, has seen patients with many ailments over the years: scabies, trench foot and, most recently, hepatitis A, which he is on a mission to vanquish. "You want a hepatitis A vaccine? There's an outbreak in Detroit," Carpenter says quietly to one older man. The man nods, rises and follows Carpenter to the conference room, where a second nurse practitioner and a team of Michigan State University medical students wait with needles.

IN BRIEF

Rising rates of hepatitis A, Legionnaires' disease and other scourges carried by viruses, bacteria and parasites are scarring U.S. cities. **Infectious diseases,** once thought to be on the wane, are making a comeback, driven by widening economic inequality and microbe-vulnerable buildings. **Not only the poor** but the well off are at risk, as disease transmission crosses lines of health and wealth.

It was January 8, 2018, and Detroit was in the midst of the largest hepatitis A outbreak in the city's history. Since August 2016 the disease had stretched across southeastern Michigan, sickening more than 770 people, and it has become the biggest such outbreak in the U.S. since a vaccine became available in 1995. Cases were still mounting as this article went to press. And Michigan's situation is not unique. Hepatitis A, caused by a virus, has infected 700 people in California—primarily San Diego—between November 2016 and the end of February, and parts of Utah, Colorado, Wyoming and Kentucky have been hit hard, too. The extent is unusual. In many other cases, the illness infects people in small, short-lived, food-associated clusters—sick restaurant workers contaminate food and infect a dozen or so customers. Sustained person-to-person spread like this is rare.

The severity of this epidemic also stands out. An estimated 81 percent of those infected in Michigan have been hospitalized subsequent to liver damage, and 25 have died, as of early March. That is much worse than the typical hospitalization rates, which hover between 11 and 22 percent, according to a 2017 report optimistically entitled *Progress Toward Viral Hepatitis Elimination in the United States,* penned by researchers at the U.S. Centers for Disease Control and Prevention. "To be making rounds on individuals who have been previously healthy but are now in liver failure is really staggering," says Stuart Gordon, a liver specialist at Henry Ford Hospital in Detroit. "We are seeing a much more severe outbreak than we've ever seen before."

Other infections have also been tearing through U.S. cities. In 2017 New York City diagnosed a record number of cases of Legionnaires' disease—65 percent more than in 2016—a serious pneumonia caused by bacteria growing in water systems. In San Francisco, rates of gonorrhea, a sexually transmitted infection that has become worryingly resistant to antibiotics, increased 22 percent between 2015 and 2016; in fact, the incidence of all three nationally reported sexually transmitted infections—chlamydia, gonorrhea and syphilis—have recently reached record highs throughout the country. The list goes on: Cases of cyclosporiasis, an intestinal illness caused by a parasite in drinking water and swimming pools, nearly doubled between 2016 and 2017. Rates of hepatitis C, another virus that damages the liver and one that often becomes chronic, have nearly tripled over the past five years.

© 2018 Scientific American

1

2

GRAFFITI in Detroit shows champion boxer Joe Louis with a homeless man called Dreadlock Mike, who died several years ago (*1*). A man enters the Tumaini Center in the city, which cares for the homeless today (*2*).

These surging infections in the U.S. are not what the medical world expected. Infectious diseases are less of a threat in this country than they were a century ago, thanks to mass vaccination, improved sanitation, and scientific advances in diagnostics, treatment and epidemiology. Rates of HIV and tuberculosis are still continuing to decline overall. But some infections are making a strong comeback in America, and researchers worry that the effects of the diseases could be more devastating now because the country has a more aged, chronically ill and vulnerable population. Infections rarely seen in the U.S. are also ripe for emergence, some scientists say, and a handful of parasitic diseases are becoming established—yet are woefully underdiagnosed. "We're going to start seeing more and more infectious disease outbreaks," says Margot Kushel, a physician and scientist at the University of California, San Francisco, Center for Vulnerable Populations, which is at the Zuckerberg San Francisco General Hospital and Trauma Center.

There are many causes for these rising infectious tides, but researchers agree that a major driver is the country's ever worsening income inequality. The disparity between America's highest and lowest earners exceeds that of virtually every other developed country, and it is still widening. The number of households earning less than $15,000 a year grew by 37 percent between 2000 and 2016. Households earning $150,000 or more increased by exactly the same amount. In poor areas, where almost half the people live below federal poverty levels, populations doubled during this period. People on these bottom rungs of society's ladder live in crowded, often unclean conditions, have limited health care, must work when sick, have poor nutrition, experience debilitating stress, and are more likely than others to abuse drugs and alcohol—all known infection risk factors.

What makes for large outbreaks, however, is that when illnesses start spreading through America's urban poor, they do not stay there. Between 2000 and 2013 the country's urban population increased by 24 million people, and crowding facilitates transmission. More city-dwelling Americans take public transportation and travel now than ever before, too, turning the nation into the equivalent of a crowded, germ-trading global market.

Changes in city infrastructure also drive up current infection risk. The huge water towers that provide buildings with

© 2018 Scientific American

1

air-conditioning are perfect breeding grounds for the bacterium that causes Legionnaires' disease; well-meaning attempts to conserve energy by reducing flow and water temperature in these tower systems "really amplify the conditions that allow *Legionella* to thrive," says Ruth Berkelman, director of the Center for Public Health Preparedness and Research at Emory University's Rollins School of Public Health. Many public housing and school buildings in U.S. cities are more than 100 years old and suffer from poor ventilation, which causes microbes to concentrate on surfaces and in indoor air pockets; some new buildings intentionally minimize ventilation to conserve energy. And water pipes are aging and increasingly leaking, breaking and becoming contaminated with microbes.

These disease-driving forces are social and economic rather than biological and medical, and they have been overlooked by many scientists and politicians. Few researchers have been studying how larger societal issues increase infection risk, and on the policy side, investment in disease prevention and control has been dropping. "We look more and more like the developing world, with very, very rich people and very, very poor people, and the very, very poor people are living in really abysmal situations," Kushel says. Inattention to this divide, and not any shortfall in medical innovation, is leaving our doors wide open for infectious catastrophe.

HOW THE OTHER HALF DIES

AN HOUR AFTER CARPENTER'S ROUNDS, a dozen men and women had received hepatitis A vaccines in the conference room. Then a shivering middle-aged African-American woman sought the help of Carpenter and his partner that day, nurse practitioner Nicole Merenius. The woman had been having chills, a cough and congestion for several days. Carpenter administered a rapid influenza test, but it came back negative. At this point, most doctors would tell their patients to go home and rest. But for this woman, a chair in the crowded Tumaini center room *was* home, at least for now. She had nowhere else to go.

The nation's homeless are among the most at risk for infectious disease for a number of reasons. They are either living on the streets, where they do not have easy access to toilets, sinks and showers, or they are staying in crowded shelters with similar problems. They are often surrounded by coughing, sneezing, sick people. Public health agencies such as the CDC tell Americans to wash their hands frequently and to stay home when ill, yet the homeless do not have the opportunity to do either. They offer profound testimony to the problem with conventional wisdom that says that infections are caused solely by germs. The truth is that a person's life and circumstances strongly shape their risk.

There are good data to back up this idea. One century ago the 1918 Spanish flu swept across the globe, taking 50 million lives.

© 2018 Scientific American

AT THE TUMAINI CENTER, people live in chairs because the center has no beds (*1*). The homeless, and the city in general, have been in the grip of a hepatitis A epidemic; at the center a man gets vaccinated against the disease (*2*).

In a 2016 study published in the *Proceedings of the National Academy of Sciences USA,* a team of biologists and epidemiologists showed how demographics and socioeconomic status changed the risk of death in Chicago during the outbreak. Although pandemics by nature are supposed to put everyone equally in peril, the researchers found something quite different. In census tracts with a higher proportion of people who were illiterate—a marker for poverty, among other things—mortality rates were much greater than in areas with high literacy rates. With every 10 percent increase in the illiteracy rate, they found, there was a corresponding 32 percent increase in flu-triggered death. Scientists discovered, too, that the flu spread much more quickly in Chicago areas that were more crowded and had higher rates of illiteracy and unemployment compared with other city regions.

The homeless might be uniquely susceptible to infectious disease, but for similar reasons home- and apartment-dwelling individuals who live in poverty are not much better off. Financial woes incite stress, which has been shown to amplify infectious disease risk. In a paper published in 1991 in the *New England Journal of Medicine,* researchers assessed the stress levels of 394 individuals and then gave them nasal drops containing either one of five types of respiratory viruses or a saline solution. The more stress the people were under, the more likely they were to fall ill if the drops they had been given contained a virus. Poor Americans also have a harder time maintaining a healthy diet. A 2017 survey of nearly 2,000 supermarkets across the U.S. found that, per serving, healthy foods such as fruits and vegetables cost twice as much as unhealthy foods such as sweets and salty snacks. Nutritious foods are also harder to find in poor areas: Johns Hopkins researchers reported that stores in lower-income Baltimore neighborhoods had less healthy food on offer than did stores in more affluent ones. And when nutritious food is hard to get, prognoses worsen. Researchers reported in a 2016 study that certain nutritional deficiencies make it more likely for patients admitted to infectious disease clinics to die.

The aging population is yet another booming disease risk. The number of Americans aged 65 and older is expected to double between now and 2060, and as individuals age, their immune systems weaken and have a harder time fighting off microbial insults. Then, when they do get sick, they fare poorly. Elderly patients are three times as likely to die from common infections than younger individuals. Pneumonia and influenza, for instance, are the fourth-leading cause of death among American adults 65 and older but only the ninth most common cause among those 25 to 44. And rates of obesity and diabetes are increasing—particularly among lower-income groups—which compounds the problem. "One of the main things that diabetes does is impair your immunity," Kushel says. "As we see more and more people living with obesity and diabetes, we're going to see more infectious diseases."

Substance abuse is also pushing susceptibility higher. The skyrocketing number of new hepatitis C infections in the U.S.—a 15-year high, according to the May 2017 CDC data—is in part because of needle sharing occurring as part of the growing opioid epidemic. Hepatitis C already kills more Americans than any other infectious disease, and the death rate is poised only to increase. The opioid epidemic could also in part explain why Michigan's hepatitis A outbreak has been so deadly: 50 percent of cases were substance users, and 27 percent of them had underlying hepatitis C, which means they were hit with two liver infections at once. Substance abuse leads to risky behavior, too, including unprotected sex. A syphilis outbreak tied to drug use swept through Oklahoma City beginning in March 2017, and during the next 12 months it sickened 241 people.

The working poor in urban areas are also uniquely positioned to spread infectious diseases because of their job conditions. More than one million low-income Americans work as food preparers, which pays an average of $13,200 a year. Many of these workers go in even when they are ill. In a 2015 study, researchers at state health departments interviewed 426 restaurant managers around the country and reported that many of the restaurants' policies regarding working while ill violated U.S. Food and Drug Administration recommendations. Seventy percent of the managers said they had worked while sick—even with a stomach bug—because they felt obligated or worried that they would not get paid otherwise. According to a 2014 report by nonprofit Families and Work Institute, only 52 percent of employers offer paid sick leave, and among those who do, 41 percent offer it only to employees who have worked there for at least a year. "You can just imagine that if people feel they have to work or they won't get paid for that time, that you're going to have a lot of sick people at work," says Jonathan Fielding, a health policy researcher at the U.C.L.A. Fielding School of Public Health. In Michigan's outbreak, 32 of the individuals who have gotten hepatitis A have been food workers; some unwittingly spread the infection because they worked while harboring the virus.

© 2018 Scientific American

Yet if the seed of an infectious illness can be traced back to poverty, when it spreads widely enough, no class gets spared. Stanley Kozlowicz, a retired General Motors engineer living in Dearborn Heights, a city 12 miles west of Detroit, needed a liver transplant after he caught hepatitis A. He believes he was infected from restaurant food in July 2017, although the county health department has not been able to pinpoint the source. "I was in really good shape, walking six and a half miles a day," Kozlowicz says. Weeks later "a doctor said, 'Well, it's in God's hands—we've done everything.'" He had his transplant in August 2017, but his health woes continued. He had a second operation because of leaking bile, a common liver transplant complication. Then he developed sepsis, and then his body partially rejected the new liver. In January 2018, Kozlowicz was back in the hospital because a liver biopsy—his third since catching hepatitis A—had caused internal bleeding. But he was optimistic. "I think we've finally turned a corner," he says.

Kozlowicz is among many middle- and upper-class Michigan residents who have fallen ill during the epidemic. "It really is a human society issue because while it may have started in [the homeless population], it can easily transmit to the whole of society—we've seen that," says Henry Ford Hospital's Gordon. One large New York City outbreak of Legionnaire's disease was traced back to the posh Opera House Hotel in the South Bronx. And sexually transmitted infections do not heed class boundaries, either. "We're seeing disproportionate increases in STDs based on social disparities," says David Harvey, executive director of the National Coalition of STD Directors, an organization representing health department STD directors and community-based partners. But "we are seeing increases across the board, and in all populations in the U.S. In fact, right now we have the highest number of STD cases in American history."

Once they reach critical mass, epidemics then become difficult to stop. Michigan's hepatitis A problem has been boiling for more than 18 months. "I don't know how long this is going to go on, but there is a possibility that despite our best efforts, this is going to run its course naturally" and continue for quite some time, says Kevin Lokar, medical director of Michigan's Macomb County Health Department, 21 miles north of Detroit. The natural course of an infection can, in some cases, take decades. In the early 1900s tuberculosis began spreading throughout Cape Town, South Africa—a city with extreme poverty as well as exorbitant wealth—and despite careful attempts to control it over the years, rates are still as high today. "Things have changed very little," said Robin Wood, director of the Desmond Tutu HIV Center at the University of Cape Town, at a conference held at Washington, D.C.'s National Academies of Sciences, Engineering, and Medicine in December 2017.

Vaccine Inequality

The same preventive shot may be less effective in poor neighborhoods than in rich ones

By Lee Riley

Health service shortfalls are often blamed for high disease rates in slums, but service problems are not the only reason poor neighborhoods fare worse than wealthy ones. Infectious diseases can differ at a basic biological level between rich and poor locales, and these differences can cripple vaccines intended to fight them.

My colleagues and I have seen these effects with rheumatic heart disease in poor parts of Brazil. This ailment has virtually disappeared from high-income countries, where antibiotics are readily available, but it is a major cause of heart trouble in less affluent nations, and it is often fatal. The disease is caused by repeated throat infections from group A streptococcal, or GAS, bacteria. When the body's immune system attacks these microbes, some proteins in heart valve cells that look similar to the bacterial proteins get attacked as well.

A vaccine against GAS could thwart such infections. But the bacteria are difficult targets. There are more than 120 different strains of these bacteria, and a typical sore throat from these pathogens can be caused by any number of these strains. Each has a different version of a gene that codes for M protein, a molecule that is a key part of the bacterium's outer membrane. To make an experimental vaccine, researchers included M proteins from 26 common strains to try to ensure immunity. Yet when scientists looked at non-European and non–North American patients with GAS infections, the 26 types appeared much less often or not at all. While these strains were frequent in high-income countries, where the vaccine was developed, they were rarities elsewhere, where the vaccine would be less effective.

Biological differences occur even within the same city. My research team compared GAS strains from children in slums and in wealthy neighborhoods in Salvador, Brazil. The collection of strains in a community is given a number called a diversity index. The greater the number of strains with different M protein genes, the higher the index. The diversity index of GAS strains of well-to-do children—they attended a private clinic and could afford private insurance—was close to that reported from high-income countries, around 0.90. But the index for slum children was higher, about 0.96. There was another distinction: the two most common strains in high-income countries accounted for 36 percent of GAS samples in wealthier Brazilian children but only for 19 percent of samples from two slum clinics. If this experimental vaccine were to be adminis-

RISKY BUILDINGS

THE WAY CITIES GET BUILT shapes infection risks, too. On March 14, 2003, a 33-year-old man from Shenzhen, China, started to feel unwell. He had a fever, muscle aches and stomach woes, but he was well enough to visit his brother on the 16th floor of block E of Amoy Gardens, a huge Hong Kong complex consisting of 19 33-story apartment towers. While he was there, the man—identified to this day only as patient YY—had a bout of diarrhea and flushed the toilet.

Ten days later other block E residents, including patient YY's brother and sister-in-law, started getting sick. By April 15, 99 people in block E had been diagnosed with severe acute respira-

© 2018 Scientific American

tered to the children in Salvador, it would be far less effective in the slums.

The higher diversity of GAS strains in slums may be the result of bacteria changing by trading genes back and forth. Trades may be easier when extreme human crowding exists, which means more frequent contact among different bacterial cells and more gene-trading chances. Strain diversification may increase the chances of a bad immune reaction against the heart.

Germ-level disparity has also been recognized as a potential problem with current widely used vaccines against cervical cancer. These injections target strains of cancer-causing human papillomavirus, or HPV. Research has indicated that a portion of African-American women living in some parts of the southeastern U.S. have different HPV strain infections, however. We do not know if these racial and geographical differences in virus types will ultimately affect vaccine effectiveness. What we do know now is that it is not just access to a clinic that determines what therapies work but also differences among disease-driving germs—distinctions created by the social environment.

***Lee Riley** heads the division of infectious diseases and vaccinology at the University of California, Berkeley, School of Public Health. He is a physician trained in epidemiology and molecular biology.*

tory syndrome, or SARS, a viral disease that kills between one and two of every 10 infected people. Another 222 people in other apartment blocks had the infection as well. The outbreak, which killed 42 people, ended up accounting for nearly 20 percent of all reported SARS cases in Hong Kong during its infamous 2002–2003 contagion.

Scientists spent months trying to understand why the disease hit Amoy Gardens so hard. In a paper published in April 2004 in the *New England Journal of Medicine,* researchers at the University of Hong Kong and the Chinese University of Hong Kong solved the mystery. Each column of bathrooms in the tall building, they discovered, was connected by vertical drainage pipes that branched out to each apartment's fixtures. Every branch had a U-shaped section of pipe that fills with water to prevent rodents, insects and sewer gases from entering each flat from the system. Some U-traps, however, were just below rarely used floor drains in each bathroom, so they never filled up. Many tenants had also installed bathroom fans that created strong negative pressure, drawing air into the bathroom from those empty U-traps. One of those traps became a viral reservoir. When patient YY, who had SARS, flushed his diarrhea on March 14 (he did the same thing during a second visit to his brother on March 19), virus particles in stool droplets made it into the trap. Then the negative pressure from the fan sucked contaminated droplets into other apartments—and also out their windows and into neighboring buildings.

So when it comes to infectious disease, buildings and infrastructure matter a lot. The design of a city determines where sewage, air and water go and whether they get contaminated along the way. If rooms are poorly ventilated, the microbes people exhale or the stool droplets that permeate the air during a toilet flush get more heavily concentrated in the air over time. Yet in an otherwise laudable attempt to promote energy conservation, restricted ventilation is now a construction trend. A handful of states, including New York, Maryland, Illinois and Massachusetts, have passed laws requiring that new homes pass stringent air-tightness tests. Older buildings are being modified, too. "I'm seeing this in college buildings and other buildings, where people have tried to conserve energy and have sealed up all the leaks," says Donald Milton, an environmental health scientist at the University of Maryland School of Public Health. In 2016 researchers at the University of Hong Kong modeled the spread of influenza in various indoor environments and concluded in a paper that "ventilation rate has a strong influence on the outbreak dynamics." Opening a window, they noted, can reduce infection risk as much as getting vaccinated.

Water pipes are another infection source, as Amoy Gardens shows, but pipe-related problems appear badly underreported. In 2013 and 2014 the CDC reported 42 disease outbreaks associated with drinking water in the U.S., but "what makes it to the CDC outbreak database is a dramatic underestimation of true waterborne disease incidence," says Kelly Reynolds, an environmental and occupational health scientist at the University of Arizona School of Public Health. A number of steps have to occur in succession for an illness to be categorized as outbreak-related: A doctor has to see a sick individual, order a lab test that comes back positive for a reportable infection and then submit the results to the CDC. Next the CDC has to decide to conduct an investigation and determine that an outbreak is indeed taking place. A more realistic estimate of true waterborne dis-

Illustration by Hanna Barczyk

© 2018 Scientific American

ease burden, Reynolds says, is that 19 million to 21 million Americans are sickened by microbe-contaminated water from taps, swimming pools, hot tubs and showers every year, based on data from epidemiological and sampling studies.

Microbes build up in water for a number of reasons. In swimming pools, hot tubs and water parks, people swim while sick or while symptom-free but still contagious, contaminating the water with infected fecal matter. (A standard dose of pool chlorine does not kill all types of germs quickly.) In addition, many drinking-water distribution pipes, particularly in older cities, have become old and leaky; estimates suggest that between 10 and 20 percent of water that leaves water utilities leaks out along the way to its destination, and "where water can leak out, contaminants can leak in," Reynolds says. Well-meaning energy conservation efforts worsen the problem by reducing water flow, which allows microbial biofilms to build up on pipe surfaces, and by reducing the upper threshold of water temperature.

The worrying increase in Legionnaires' disease in the U.S. highlights yet another water-related challenge. Cooling towers are increasingly popular in urban areas because they efficiently cool large buildings through water-based heat exchange. But *Legionella* bacteria, which naturally occur in water, thrive in such warm conditions and can sicken people when they get aerosolized and pumped out of air vents. Aerosolized shower water, fountain water and even supermarket vegetable misters can pose risks, too. Although some of the increase in Legionnaires' diagnoses—which rose by a factor of five in the U.S. between 2000 and 2015—could be the result of increased awareness and testing by doctors, scientists argue that the infection is likely becoming more common. "One hundred years ago we didn't have heating ventilation and cooling systems the way we do now. Now we've got these bigger buildings, we've got the increased complexity of these plumbing systems," Emory's Berkelman says.

Large cities also increasingly struggle with waste management—and where there is waste, there are disease-carrying rodents. A 2007 study reported that 65 percent of rats tested in Baltimore were infected with leptospirosis, a bacterial disease that people can catch from rat urine. It can cause renal failure and lung hemorrhage. (Pets are also at increased risk, so much so that there is now a popular canine leptospirosis vaccine.) It is another potentially underestimated source of disease; some scientists worry that it is somewhat common but treated by doctors as an unidentified bacterial infection. "There's a laundry list of pathogens that infect both humans and animals that you find in urban rats," explains Gregory Glass, an infectious disease researcher at the University of Florida's Emerging Pathogens Institute. "Yet if you ask how many cases of any of those have been spotted in the past 10 years, the answer is probably pretty close to zero, not because that's the real background but because people just don't look for it."

Rarely diagnosed infections known as neglected tropical diseases are also likely to be more common than doctors expect. These include Chagas disease, a blood-borne parasitic infection transmitted through the bite of a kissing bug. CDC researchers estimated in a 2009 paper that more than 300,000 Americans suffer from Chagas and that 30,000 to 45,000 of them develop heart disease or heart failure every year as a direct result. The agency also estimates that 1.1 million are infected annually with trichomoniasis, a parasitic sexually transmitted disease, and that 1,000 are hospitalized yearly with neurocysticercosis, a brain tapeworm that causes epileptic seizures. "Most physicians are poorly trained in parasitic and tropical diseases. They don't realize they're widespread," says Peter Hotez, dean of the National School of Tropical Medicine at the Baylor College of Medicine. "And they're mostly diseases of people who live in poverty in the U.S.—that's another reason they don't get attention."

SOCIAL VULNERABILITY

SCIENTISTS AND PUBLIC HEALTH AGENCIES are starting to acknowledge that social factors such as poverty and living environment play an enormous role in infectious disease, yet little research directly investigates the link between the two. "It's kind of a macro-level factor that is surely behind the risk, but it's not specifically examined in much of the research," says Stephen Hwang, director of the Center for Urban Health Solutions at St. Michael's Hospital in Toronto.

One reason for this inattention is lack of money. The National Institutes of Health likes to fund research that is focused on specific diseases—on the epidemiology of hepatitis A, for instance, rather than the relation between homelessness and hepatitis A—so a broader exploration of these links "doesn't lend itself to a sustainable research career," Hwang says. Hotez agrees, noting that the study of the social causes of disease requires interdisciplinary investigations, work that is not generally well supported. "We don't have a good mechanism in the U.S. to bridge disciplines," he says. "We don't think about teaming up with an economist, political scientist and anthropologist to solve these problems. But we're paying the price for that."

The Michigan hepatitis A outbreak is a perfect example of this type of tunnel vision. To curb the infection surge, the state is pouring resources into vaccine distribution, a disease-centric approach that will undoubtedly help but also overlooks the many problems, such as substance use and dangerous sick-leave policies, that made it possible for the outbreak to occur in the first place. (Many low-income Michigan residents are also refusing the vaccine, which health department officials suspect could be caused by distrust of the medical and political establishment.) And what if there is no vaccine for the next infection? In recent years pharmaceutical companies have indicated a waning interest in vaccine development; emergency vaccine efforts to combat sudden epidemics are especially expensive, risky ventures. "It's very disruptive, and that's not the way that we want to do business going forward," Rip Ballou, director of the U.S. research and development center for GSK Global Vaccines, told STAT News in January. Novartis closed its vaccine division in 2014.

When the U.S. does put money into controlling an infectious disease, it also tends to stop once things improve, says Ron Valdiserri, a senior research associate at the Johns Hopkins Bloomberg School of Public Health and former deputy assistant secretary for health responsible for infectious diseases at the U.S. Department of Health and Human Services. There is "this notion that 'Oh, once we have a good handle on disease, we can forget it and move on to something else,'" he says. But "with many of these infectious diseases, even when you're successful and you can reduce new infections and you reduce incidence, they can spring back up again." HIV is following this pattern. Although overall incidence in the U.S. has been dropping, in some poor urban areas, ethnic groups and areas of the South, the opposite is true.

© 2018 Scientific American

TRINITY CEMETERY stretches in front of the abandoned giant Packard automotive plant in Detroit, which provided jobs and a decent living for thousands before it closed in the mid-1950s.

Although funding amounts will ultimately be decided by Congress, the Trump administration's budget request for the 2019 fiscal year for the CDC slashes $43 million from current programs for STD and tuberculosis prevention. (This includes HIV and viral hepatitis programs.) It cuts $704 million from public health preparedness and response, $44 million from immunization and respiratory diseases, and $60 million from emerging and zoonotic diseases. State and local health departments, which investigate and control infectious disease outbreaks on the ground as they occur, are also suffering. And their problems cannot be laid at the door of the current administration. In April 2016, before Donald Trump won the Republican presidential nomination, Gail Bolan, director of the CDC's Division of STD Prevention, noted in a congressional briefing that 43 percent of state and local health department STD clinics had reduced their hours, and 7 percent had closed their STD clinics entirely because of funding cuts.

Back at the windowless Tumaini Center, Carpenter cares for as many people as he can. He hands out medicine, answers questions, offers snacks, asks residents how they are doing. He only works there three days a week, so addressing their many health needs is a constant challenge. To make matters worse, financial support for the Tumaini Center's parent organization, the Neighborhood Service Organization, which provides programs and services for at-risk Detroit residents, is waning. From 2013 to 2016 the organization's revenue from donations and grants dropped by more than 20 percent.

The U.S. has come a long way since its early battles with smallpox, cholera and polio. But modern medicine isn't a panacea. The lives of microbes, like those of people they afflict, are shaped by their environments—and those environments are closely interwoven. As the country takes resources away from vulnerable citizens, it unwittingly enriches the strength of plagues among them. SA

Editors' note: *Since this article was published in our May 2018 issue, the hepatitis A outbreak has spread more widely to states such as West Virginia and Kentucky. The latter had 1,788 cases between November 2017 and September 2018. The outbreak is specifically linked to drug use—particularly injectable opioids—and homelessness.*

Contributing editor **Melinda Wenner Moyer** won an Award for Excellence in Health Care Journalism for her December 2016 *Scientific American* article "The Looming Threat of Factory-Farm Superbugs."

MORE TO EXPLORE

Environmental Transmission of SARS at Amoy Gardens. Kelly R. McKinney et al. in *Journal of Environmental Health*, Vol. 68, No. 9, pages 26–30; May 2006.

Global Rise in Human Infectious Disease Outbreaks. Katherine F. Smith et al. in *Journal of the Royal Society Interface*, Vol. 11, No. 101, Article No. 20140950; December 6, 2014.

The State of the Nation's Housing 2016. Joint Center for Housing Studies of Harvard University. President and Fellows of Harvard College, 2016.

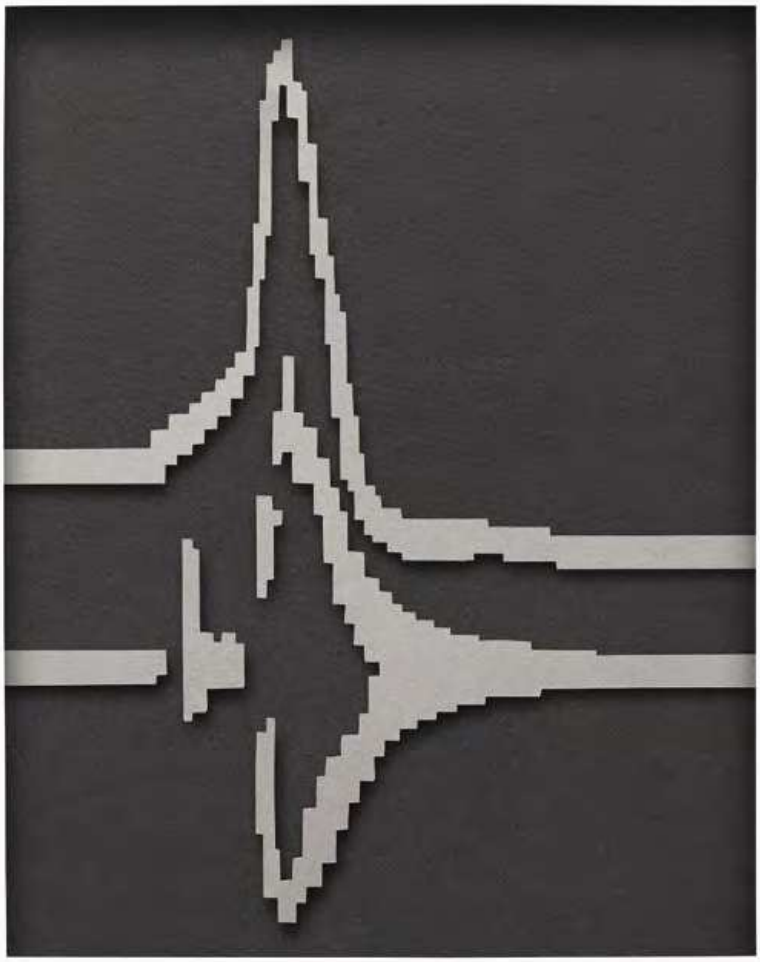
Potential Impact of a Ventilation Intervention for Influenza in the Context of a Dense Indoor Contact Network in Hong Kong. Xiaolei Gao et al. in *Science of the Total Environment*, Vols. 569–570, pages 373–381; November 1, 2016.

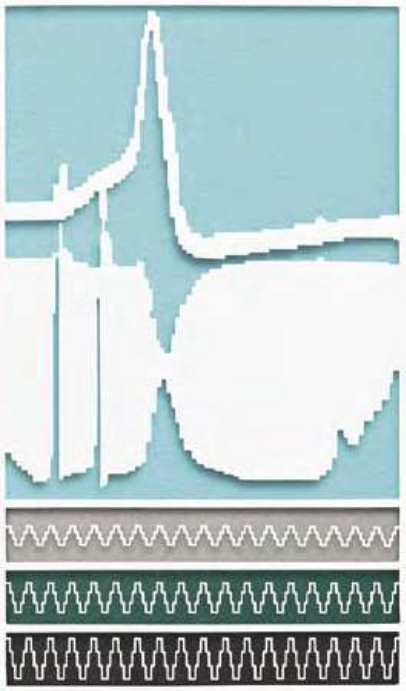
FROM OUR ARCHIVES

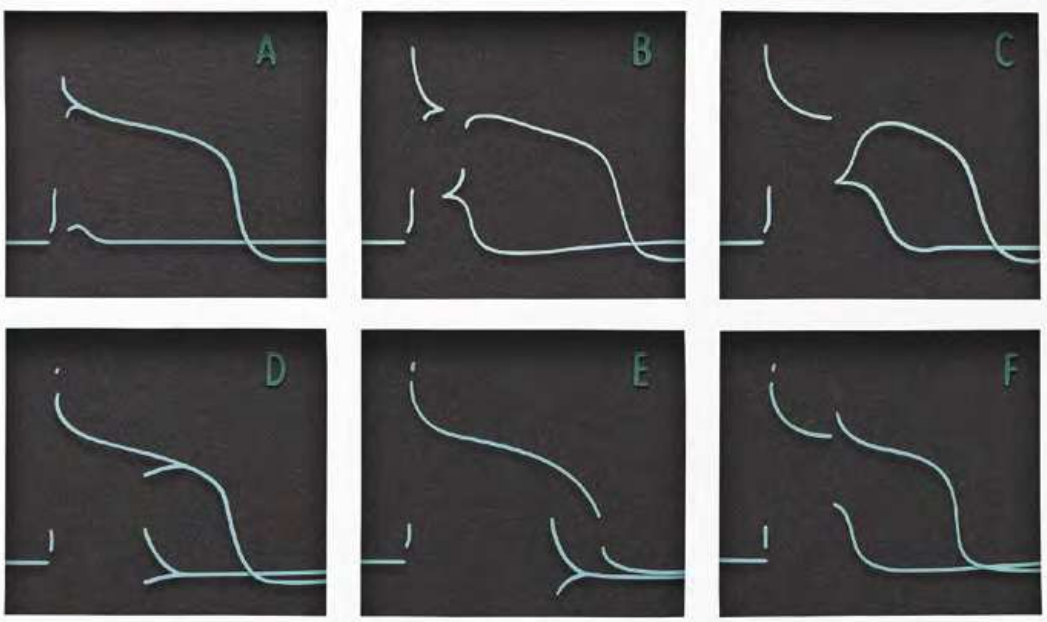
Sick of Poverty. Robert Sapolsky; December 2005.

scientificamerican.com/magazine/sa

© 2018 Scientific American







© 2018 Scientific American

NEUROSCIENCE

THE BRAIN, Reimagined

Physicists who have revived experiments from 50 years ago say nerve cells communicate with mechanical pulses, not electric ones

By Douglas Fox

Illustration by Owen Gildersleeve

© 2018 Scientific American

A YOUNG WOMAN WITH WAVY BROWN HAIR AND MAROON NAILS lay on a gurney in a hospital room in Copenhagen. Her extended left arm was wired with electrodes. A pop pierced the air every few seconds—an electric shock. Each time, the woman's fingers twitched. She winced. She was to receive hundreds of shocks that day.

The woman, attended to by several physicians in laboratory coats, was renting out her arm for 1,000 Danish kroner, about $187. Thomas Heimburg, a physicist trained in quantum mechanics and biophysics, sat on a stool, safely out of the way, sketching on his iPad the details of a harsh experiment that he hoped would produce profound results.

The physicians had injected the woman's arm with the anesthetic lidocaine—a dose strong enough to deaden her limb for surgery. At first, the nerves in her arm did not respond to the shocks. But the attendants gradually dialed up the current. At this moment, the jolts were 40 milliamperes, nearly 10 times their original strength—similar to the electricity coursing through a five-watt lightbulb.

Pop—another shock. The woman's hand twitched like a dying snake. Heimburg paid no notice as he stared at a computer monitor on the wall. A waveform depicting the electric signal in the arm muscle and nerve leaped across the screen in one large spike—evidence that the ever increasing shocks had started to overcome the anesthetic. The nerve was now firing as strongly as it did before the woman was anesthetized. Heimburg was pleased. "The things that are written in books," he said quietly, "they are in contradiction to this."

Heimburg, who works at the Niels Bohr Institute in Copenhagen, famous for physics research, hopes to contradict lots of things written in books. This experiment, which I witnessed in December 2011, was designed to investigate a long-standing medical mystery.

Physicians have administered general anesthetics for 170 years. They have discovered dozens of effective compounds. When given at progressively higher doses, the drugs all silence nerve functions in the body and brain in the same distinct order: first memory formation, then pain sensation, then consciousness, and eventually breathing. This same sequence happens across all animals, from humans to flies.

Yet no one knows how anesthesia actually works. The molecular structures of nitrous oxide, ether, sevoflurane and xenon are so different that it is unlikely they exert their common effects by binding to equivalent proteins in cells, as other drugs do.

Heimburg thinks anesthetics work in a radically different way: by changing the mechanical properties of a nerve. If that is true, it means that nerve cells, or neurons, throughout the body and brain are mechanical machines, not the electric circuits scientists have believed in for decades. In Heimburg's view, the electric pulses are simply the side effects of a physical shock wave that ripples down the nerve, similar to the way sound waves travel. He thinks anesthetics silence nerves by soaking into the fatty membranes that encase nerve fibers, rendering them too soft to transmit the shock waves, like a guitar string too slack to twang.

It was tempting to dismiss Heimburg as nutty when I watched that experiment. But in the seven years since then, he and his colleagues have rolled out an array of evidence: delicate measurements of how mechanical waves move through single

IN BRIEF

Physicist Thomas Heimburg may upend the biology world. He says nerves do not fire electrically but mechanically. To prove it, he is reviving experiments done 50 years ago by a discounted neuroscientist.

Heimburg's radical idea is that a signal traveling down a pipelike nerve fiber is a compression wave, like sound, which temporarily changes the fatty membrane that coats the pipe from fluid to crystalline.

Biologists think Heimburg is revealing only side effects of an electric pulse, yet a few are acknowledging the two actions might work in concert, which would transform explanations about how the brain functions.

© 2018 Scientific American

nerve cells and of how much and how quickly the membranes can expand and contract, as well as studies showing how anesthetics alter these properties. Other scientists are starting to take an interest. Now Heimburg is preparing for a crucial experiment that could clinch his case: measuring the heat emitted by a single nerve cell as a pulse shoots through it.

Heimburg's work continues to demonstrate that the nerve pulse is more complex than most biologists may realize. The mechanical components may have been overlooked because of an accident of history: 50 years ago off-the-shelf instruments could readily measure the tiny electric impulses in neurons but not the mechanical ones. Hardware limitations influenced which discoveries scientists made and which ideas entered mainstream scientific thought. Heimburg's experiments are now reopening a decades-old scientific schism.

The story of the mechanical neuron holds lessons for all of science about biases and accidents of history. It also could change our basic understanding of nerves, brains and intelligence. Scientists have struggled to explain how brains achieve such daunting feats as face recognition and conversation while relying on proteins in neurons that are electrically noisy and unreliable. Heimburg is showing how the mechanical waves may compensate for this noise. If his theory proves out, he could rewrite biology. Or he might just be wrong.

HOT NERVES

THE NEURAL PULSE that scientists have tried for so long to explain lasts for only an instant. Step on a thumbtack, and your brain senses the pain within a fraction of a second. The signal travels through nerve fibers at up to 30 meters per second.

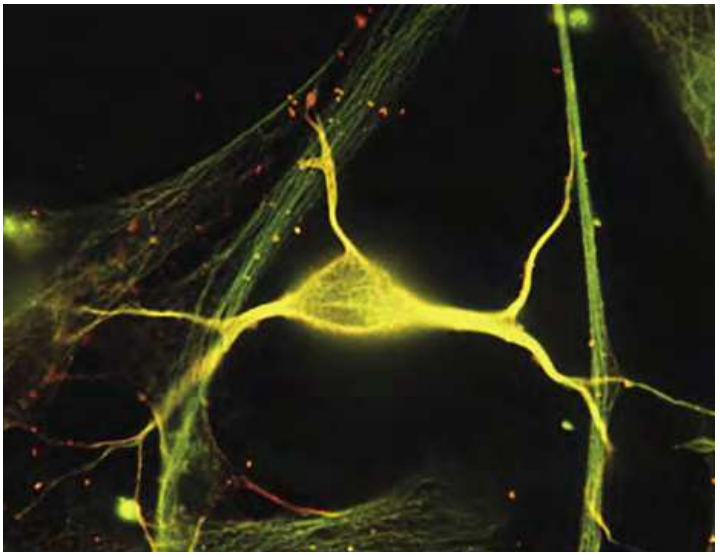
The fibers resemble tiny hollow pipes, finer than a hair. The pipe wall is formed by an oily cell membrane. Charged sodium and potassium atoms, called ions, hover around the inside and outside of the membrane. By the mid-1900s researchers had learned to stick electrodes into nerve cells to monitor the voltage across the membrane wall. They discovered that as a nerve pulse travels down the membrane and passes the electrode, the voltage spikes for several thousandths of a second. In 1952 two British scientists, Alan Hodgkin and Andrew Huxley, reported that the spike happens as sodium ions stream through the membrane wall from outside to inside. The voltage then reverses to normal as potassium ions gush through the membrane from inside to outside. The Hodgkin-Huxley model became the foundation of modern neurophysiology.

Hodgkin and Huxley received a Nobel Prize in 1963. But a few scientists continued to unearth observations that undermined their model, observations that Heimburg has re-created, even though some of those scientists had been written off as misguided.

Ichiji Tasaki, a senior neurobiologist at the National Institutes of Health for many years, was one of them. In 1979 he conducted an unorthodox experiment. Gazing through a microscope, he gingerly placed a fleck of shiny platinum atop a fine white thread—a nerve fiber bundle of a crab, laid bare by dissecting the animal's leg—and trained a laser onto the platinum. By measuring the reflection of the laser light, he could detect motions that would show whether the nerve bundle briefly widened or narrowed as an electric pulse passed by. He and his then postdoctoral fellow, Kunihiko Iwasa, took hundreds of measurements. After a week, the answer was clear: every time a pulse shot through the nerve fibers, they briefly widened, then narrowed again, within a few thousandths of a second.

The ripple was minuscule: the membrane surface rose by only about seven billionths of a meter. But it coincided perfectly with the passing electric pulse, confirming a suspicion Tasaki had harbored for years: that Hodgkin and Huxley were wrong.

As far back as the 1940s, researchers had noticed that as an electric pulse passes through a nerve fiber, the translucent cell briefly becomes more opaque. By 1968 Tasaki and another team found evidence suggesting that as the pulse arrives, molecules in the membrane physically rearrange themselves, then revert to their original configuration after the pulse passes.



NEURON, or nerve cell (*yellow*), in the brain's hippocampus—the center of long-term memory—is supported by proteins (*green* and *red*).

Then there was the heat. Researchers expected an electric pulse to release heat—common when electricity flows. But several teams discovered something strange. A nerve fiber's temperature rose several millionths of a degree Celsius as a pulse raced by, yet after it passed, the temperature quickly fell again. The heat had not dissipated; instead the nerve had reabsorbed most of it, also within a few thousandths of a second. For Tasaki, the transient widening, the rearranging molecules, and the heating and cooling pointed to a startling conclusion: the nerve signal was not just a voltage pulse; it was every bit as much a mechanical pulse. Scientists who listened to nerves with electrodes were missing much of the action.

Tasaki would spend the rest of his life probing these effects. He came to believe that they originated not in the cell membrane but in a layer of protein and carbohydrate filaments just underneath it. According to his theory, as the voltage pulse

ROBERT S. McNEIL *Science Source*

© 2018 Scientific American

arrives, the filaments absorb potassium ions and water—causing them to swell and warm—a process that then reverses itself after the pulse passes by.

As Tasaki pursued these ideas, he gradually fell out of step with the field. Other factors conspired against him. Having grown up in Japan, he spoke stilted English. "You [had] to know a lot of things to have a really substantive conversation with him," says Peter Basser, an NIH section head in neuroscience who knew Tasaki for 20 years. "And I think a lot of people thought he wasn't really as deep and perceptive as he was." And although Tasaki collaborated with visiting scientists, he did not produce student protégés who would carry his ideas forward.

Emblematic of the schism was the ideological rivalry that arose between Tasaki and another prominent NIH neuroscientist, Kenneth Cole, who adhered to the mainstream view. Although the two men occupied the same lab building from the 1950s to the 1970s, they barely spoke for 15 years, except at public presentations, where one would undermine the other by standing up in the audience and posing prickly questions.

Tasaki gave up his lab during an NIH reorganization in 1997 and moved into a small space in Basser's lab. He continued working seven days a week, well into his 90s. One day in December 2008, as he walked near his home, he lost his balance and banged his head on the ground. He died a week later at the age of 98.

By then, Tasaki's work had disappeared from sight. "I don't think anybody disputed that those things were being seen, because he was respected in the lab," said Adrian Parsegian, a biophysicist at the University of Massachusetts Amherst, who was at the NIH from 1967 to 2009. Rather Tasaki's findings "were explained away as not central" to nerve signaling—nothing more than side effects of the voltage pulse. The underlying scientific questions "didn't get resolved," he said. "One side got into the textbooks, and the other one didn't."

FATTY LIQUID BECOMES CRYSTAL

HEIMBURG CAME ACROSS Tasaki's work in the mid-1980s, while pursuing his Ph.D. at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. Soon he found himself immersed in long sessions at the library, poring over old papers. He would eventually connect the dots in a different way than Tasaki had. He believed that the mechanical wave, the optical changes and the transient heat must occur in the fatty cell membrane of nerves throughout the body and brain, not in the protein and carbohydrate filaments below the membrane, as Tasaki had thought.

By the late 1990s Heimburg had begun doing his own experiments, compressing artificial cell membranes to see how they might respond to a mechanical shock wave. This work revealed something crucial: the membrane's oily lipid molecules are normally fluid and randomly oriented, but they hover close to what chemists call a phase transition. Squeeze the membrane just a little bit, and the lipids condense into a highly aligned liquid crystal.

These experiments led Heimburg to declare that a nerve pulse is a mechanical shock wave that travels down the nerve membrane. As it advances, it should squeeze the membrane's lipid molecules into a liquid crystal—a phase change that would release a small amount of heat, just as water does when it freezes. Then, as the tail end of the shock wave passes, a few thousandths of a second later, the membrane would revert to a fluid state, reabsorbing the heat. That brief transition into a liquid crystal and back would also cause the nerve membrane to widen briefly, just as Tasaki and Iwasa had seen when they shined a laser on that platinum fleck.

Heimburg's experiments went one crucial step further. They showed how the shock wave and phase transition might be linked to the voltage spike that occurs as the pulse passes by. Heimburg found that he could push a membrane into its liquid-crystal state simply by putting it under a voltage. "People applied voltage across biologic membranes for 70 years or so, and none of these electrophysiologists had ever checked" for a liquid-crystal structure, he said.

Textbook diagrams portray cell membranes as thin, passive sheets of insulation wrapped around pipelike nerve fibers. But physicists are starting to realize that cell membranes have surprising properties. They belong to a class of materials known as piezoelectrics, which can convert mechanical forces into electric forces, and vice versa. Quartz watches run on this principle. This means that a voltage pulse traveling down a membrane will carry with it a mechanical wave. And conversely, a mechanical wave traveling down a membrane will express itself as a voltage pulse.

When Heimburg and his fellow researcher Andrew D. Jackson first published the theory in 2005, they had still never observed one of these electromechanical pulses in motion. One of Heimburg's former students filled that gap. In 2009 Matthias Schneider, a biophysicist now at the Technical University of Dortmund in Germany, reported that he could trigger a mechanical wave by applying a voltage pulse to an artificial membrane. The pulse strength was similar to that found in nerve cells. The shock wave traveled at approximately 50 meters per second, similar to the speed at which thumbtack-triggered signals race from the foot to the brain. By 2012 Schneider had confirmed that the mechanical and voltage pulses were part of the same membrane wave.

Schneider's most important finding came in 2014, however. A key feature of a nerve pulse is that it is all-or-nothing. If a neuron receives a weak incoming shock, it will not fire a voltage pulse. If the shock is strong enough, it will fire. "There is a threshold," Schneider says. He found that the electromechanical waves on his artificial membranes were indeed all-or-nothing. The determining factor seemed to be whether the membrane was squished hard enough to force it into liquid-crystal form. Only then, he says, "you get a pulse."

ANESTHESIA EXPLAINED

WHY HAD HEIMBURG first committed to this view of nerves and anesthesia? Hoping to find out, I visited him at his office at the Niels Bohr Institute during the same week I witnessed the hospital experiment.

Heimburg had the bookshelves of a physicist, not a biologist, crammed with volumes by dead German physicists. Among them was a row of clothbound books by Hermann von Helmholtz, who in the mid-1800s formulated a key premise of thermodynamics, that energy can change form but cannot be created or destroyed. Helmholtz, incidentally, also measured the speed of nerve pulses. "I find it absolutely mandatory to read these old texts," Heimburg said. They document the gradual

© 2018 Scientific American

How Do Nerves Send Signals?

For decades scientists have endorsed a standard explanation for how nerve cells (neurons) transmit signals in the body and brain: each message is carried as an electric impulse that travels down a cell's long axon, jumping to the next neuron. But now a handful of physicists, who have performed exotic tests on the cells in action, say the signal is actually a mechanical pulse that ripples down the axon—akin to a sound wave or seismic wave. Some researchers say any physical pulse is just a side effect of the electric impulse. Settling these contentions could revise explanations of how the brain works.

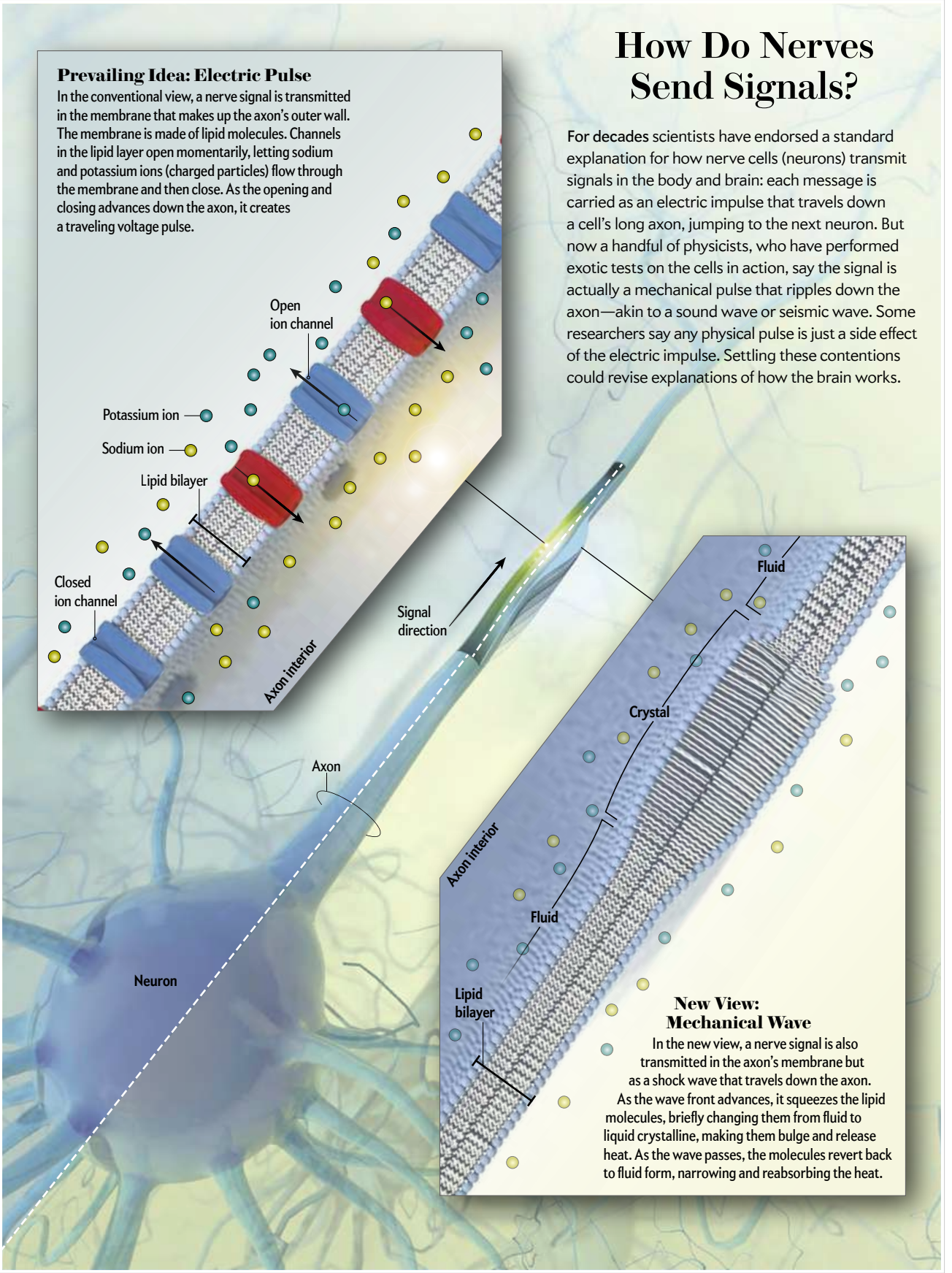

Prevailing Idea: Electric Pulse
In the conventional view, a nerve signal is transmitted in the membrane that makes up the axon's outer wall. The membrane is made of lipid molecules. Channels in the lipid layer open momentarily, letting sodium and potassium ions (charged particles) flow through the membrane and then close. As the opening and closing advances down the axon, it creates a traveling voltage pulse.
Open ion channel
Potassium ion
Sodium ion
Lipid bilayer
Closed ion channel
Axon interior
Signal direction
Axon
Neuron
Fluid
Crystal
Axon interior
Fluid
Lipid bilayer
New View: Mechanical Wave
In the new view, a nerve signal is also transmitted in the axon's membrane but as a shock wave that travels down the axon. As the wave front advances, it squeezes the lipid molecules, briefly changing them from fluid to liquid crystalline, making them bulge and release heat. As the wave passes, the molecules revert back to fluid form, narrowing and reabsorbing the heat.

Illustration by Falconieri Visuals

© 2018 Scientific American

discovery of fundamental connections among energy, temperature, pressure, voltage and phase transitions. These principles underlie Heimburg's ideas about nerve function, the ideas of a physicist pushing his way into another field. "Thermodynamics is the most profound science that we have," he said. "If you know thermodynamics, you are wise."

He was quick to point out weaknesses in popular explanations about anesthesia. Biologists think anesthetics silence nerves by binding to and thus blocking ion channels—valves in a nerve membrane that open and close to allow sodium or potassium ions to flow through. Biologists say the flow of ions propels voltage pulses down a nerve fiber—commonly portrayed as an electric signal. But because different anesthetics have vastly different molecular structures, Heimburg could not believe they could all bind to ion channels. That explanation was "completely ridiculous," he said, with a hint of frustration, as if pointing out something that should be obvious. Something "deeper, more profound," must be at work.

Heimburg's ideas were shaped in part by an old volume entitled *Studien über die Narkose,* or *Studies of Narcosis,* published by Ernest Overton in 1901. It recounts a particular experiment that caught his attention. Overton took dozens of different anesthetics and put each into a flask of water with a layer of olive oil floating on top. He shook each flask, then waited for the water and oil to separate again. He measured how much of each drug ended up in the oil versus the water. The more potent an anesthetic was in animals, the more strongly it moved into the oil, a striking result later confirmed for modern anesthetics. Olive oil and cell membranes are composed of the same oily molecules, called fatty acids. Heimburg surmised that the drugs might work by soaking into the cell membranes, altering their physical properties.

Experiments with synthetic membranes support that idea. When Heimburg infuses a membrane with an anesthetic, it prevents the membrane from becoming a liquid crystal. It does so by lowering the temperature (and raising the pressure) at which the phase transition from fluid lipid to crystalline lipid occurs—just as salt or sugar lowers the freezing point of water.

Heimburg reasoned that preventing this transition in a membrane would stop a mechanical pulse from advancing down a nerve fiber, explaining why anesthetics deaden nerves. And notably, he predicted it should be possible to overcome this effect. To create higher pressure to solidify a membrane using an electric shock, you have to crank up the current—exactly what the physicians did to the woman's arm at the hospital in Copenhagen. Stronger electric shocks did indeed overcome the anesthetic. If anesthesia can be overcome by pushing harder on a membrane with electricity, then it should also be reversible by increasing the physical pressure on a membrane.

Biologists demonstrated this way back in 1942. They used two different anesthetics, ethanol and urethane, to inebriate tadpoles to the point that they stopped swimming. Then the scientists put the animals in a hyperbaric chamber and raised the pressure to 136 times that of the atmosphere. The anesthetic effect vanished: the tadpoles resumed swimming. When the pressure was lowered, the animals again fell motionless. "It's very surprising," Heimburg said, with a smile. "How would you have the idea to put drunken tadpoles under pressure?"

THE EXISTENCE OF MECHANICAL WAVES IS NOT IN DOUBT, SAYS ONE NEUROSCIENTIST IN THE MIDDLE OF THE CONTROVERSY: "THE QUESTION IS WHETHER NEURONS ACTUALLY USE THEM TO DO SOMETHING USEFUL."

NO TOLERANCE FOR DEBATE

TO THIS DAY, Heimburg is frustrated by the way biologists react to his ideas, which he calls soliton theory (a soliton is a self-sustaining wave that maintains its shape as it travels). He has faced opposition from the moment he published his theory in 2005 in the *Proceedings of the National Academy of Sciences USA,* despite that journal's high regard.

One critic, Catherine Morris, a prominent neurobiologist emeritus at the Ottawa Hospital Research Institute, told me that the whole line of work reeks of superiority from a physicist who thinks he can simply march into a different field and set people straight. She summed this up in a favorite witticism of hers: "It strikes me as this business that physicists do, saying, 'We can approximate this cow as a single point.'"

To some extent, Morris's reaction is understandable. It is one thing to say that nerves are mechanical as well as electrical. It is quite another to reject the concept that ion channels play a role in nerve conduction—which Heimburg and Schneider do, in their biggest and most problematic departure from mainstream biology. Never mind that scientists have discovered hundreds of ion channel proteins. Or that the ion flows can be selectively altered with drugs. Or that mutations scientists can create in the proteins change the way neurons fire. "They just blithely ignore vast amounts of biology," says Morris, who spent 30 years studying ion channel proteins.

Heimburg and Schneider acknowledge that these proteins must serve some function. But they point to experiments, some by Heimburg, showing that ions can flow across artificial membranes even without channel proteins. They attribute this flow to transient holes that appear as the membrane shifts between fluid and liquid-crystal phases, and they think it happens in nerves in the body and brain.

Their skepticism reflects a cultural tendency in physics: a belief that all things should be explainable through thermodynamic principles. Biologists, they say, have neglected these principles as they fixate on proteins. A similar brand of puritanism may have facilitated the eventual dismissal of Tasaki's theory. He "did not like the term 'ion channels,'" said former postdoc Iwasa when we spoke in late 2017. This iconoclastic outlook may have guided Tasaki to discover things that others could not have, Iwasa said, "but later on, it may not have helped" him.

Brian Salzberg agrees. He studies nerve physics at the University of Pennsylvania and began his neuroscience career in

© 2018 Scientific American

1971, crossing paths occasionally with Tasaki. "He was a very clever experimenter, and I have no doubt that he measured real changes" in nerve thickness, Salzberg said earlier this year. "But he misinterpreted them." Salzberg says nerve fibers temporarily swell as a voltage pulse goes by in part because water molecules flow into the membrane through the same ion channels that let in sodium and then flow back out through the ion channels that let out potassium. If Tasaki had accepted the idea of ion channels, he might have been open to other interpretations of the mechanical wave.

But another powerful factor may have helped push Tasaki out of sight—holding an important lesson for all of science today.

IDEOLOGUES

IT IS INTRIGUING that the thermal energy of a firing nerve may be twice as large as the energy in the electric signal that has dominated neuroscience. The fact that these nonelectric features fell out of favor may stem, in part, from a quirk of history.

Tasaki was a gifted instrument builder who cut his scientific teeth in Tokyo during World War II. Faced with severe equipment shortages, he assembled his own instruments from stray electric components. Years later in the U.S., he used these skills to build exquisite, one-off instruments that measured the heat, or temporary expansion, of nerve cells.

Those devices, and expertise, never found their way to other scientists. Measuring the electric nerve signal was different. Scientists created easily transferable methods, such as inserting a tiny electrode into a cell membrane. As these techniques spread from one lab to another, so did the electrical view of nerve signaling. "There's a cultural bias," Parsegian admitted. "People look with a tool that they feel they understand, and they don't use one that they don't understand. It could have tilted the thinking."

Today the technical gaps are starting to disappear. As I checked in with Heimburg between 2011 and 2018, he gradually repeated one old experiment after another, using modern technologies to clarify the surprising things that Tasaki and others first saw decades ago. In 2014 Heimburg redid the drunken-tadpole experiment, using synthetic membranes instead of animals: as he cranked the pressure up to 160 atmospheres, the impacts of anesthetics were reversed—except that this time, Heimburg could link the effect directly to phase changes in the membrane. In 2016 he used microscopy to precisely measure, in a single cell, the mechanical wave that Tasaki and Iwasa first documented in 1979.

Heimburg, now 58, is seeking funding for what could be the most critical experiment of all: measuring the heat as a nerve pulse, or action potential, passes by. Tasaki had measured heat from bundles of fibers, but Heimburg plans to use a microchip that will measure the heat blip of a single neuron. This experiment could address a key criticism of his theory: that a nerve membrane's brief phase change from liquid to crystal should release, and reabsorb, more heat than Tasaki ever saw. Heimburg contends that the old experiments systematically underestimated the heat; because they measured many neurons, the heat reabsorption after early pulses canceled out the heat releases of later pulses. "The true signal is probably much higher," he told me in late 2017. If his measurements bear out, they could bolster his claim that the membrane transmits a mechanical wave.

Perhaps most significantly, other scientists are stepping in—outsiders who are not polarized by the old, calcified disputes. Nongjian Tao, a biosensor engineer at Arizona State University, is using lasers to track mechanical pulses in single nerve cells—like Tasaki and Iwasa did, except that Tao reflects his light directly off the nerve rather than a tiny platinum mirror, making the measurement more sensitive. He hopes to monitor hundreds of individual neurons in nerve networks at once, with lasers sensing mechanical waves as they ripple to-and-fro. Such work could answer a key question. "The existence of these [mechanical] effects is not in doubt," says Simon Laughlin, a neuroscientist at the University of Cambridge. "The question is whether neurons actually use them to do something useful."

Laughlin does not work on mechanical waves, but as someone who has studied ion channels for 45 years, he imagines that the waves could influence the little protein valves. Recent experiments show that the valves are extremely sensitive to mechanical forces in the membrane. If mechanical waves help to open and close ion channels, that could profoundly change our understanding of the brain because firing neurons mediate all thinking. Ion channels are notoriously noisy and jittery: even tiny thermal vibrations can cause them to pop open or close randomly. Information theorists have struggled for decades to explain how the brain can achieve reliable cognition using such unreliable channels. But mechanical waves could mean the openings and closings are purposeful. "That's a definite possibility," Laughlin says.

There are hints that this could be true. Some neurons in the mammalian cortex seem to violate the Hodgkin-Huxley theory. When they fire at high rates, their ion channels open more quickly, as a group, than expected. One explanation is that the channels are responding en masse to a sudden change in the membrane—the arrival of a mechanical wave that opens them more or less in unison, allowing them to fire faster than they otherwise could. The speed might allow them to transmit information at phenomenally quick rates—a possible basis for cognition. In this view, a nerve pulse is both electrical and mechanical.

Heimburg and Schneider occupy a strange place in all of this. They could perhaps one day share a Nobel Prize. Or they could end up nowhere, transfixed by the same insistence that gripped Tasaki for so many decades. The fact that some neuroscientists such as Laughlin and other experts such as Tao are interested in mechanical waves would seem like an important opening for the physicists. But Heimburg was steadfast when we spoke in early 2018. "What many people try to do is somehow rescue the Hodgkin-Huxley model by just combining it with the view that we have," he said. "But I personally … would not accept any kind of compromise between the two models." SA

Douglas Fox writes about neuroscience and extreme climates from California.

MORE TO EXPLORE

On Soliton Propagation in Biomembranes and Nerves. Thomas Heimburg and Andrew D. Jackson in *Proceedings of the National Academy of Sciences USA*, Vol. 102, No. 28, pages 9790–9795; July 12, 2005.

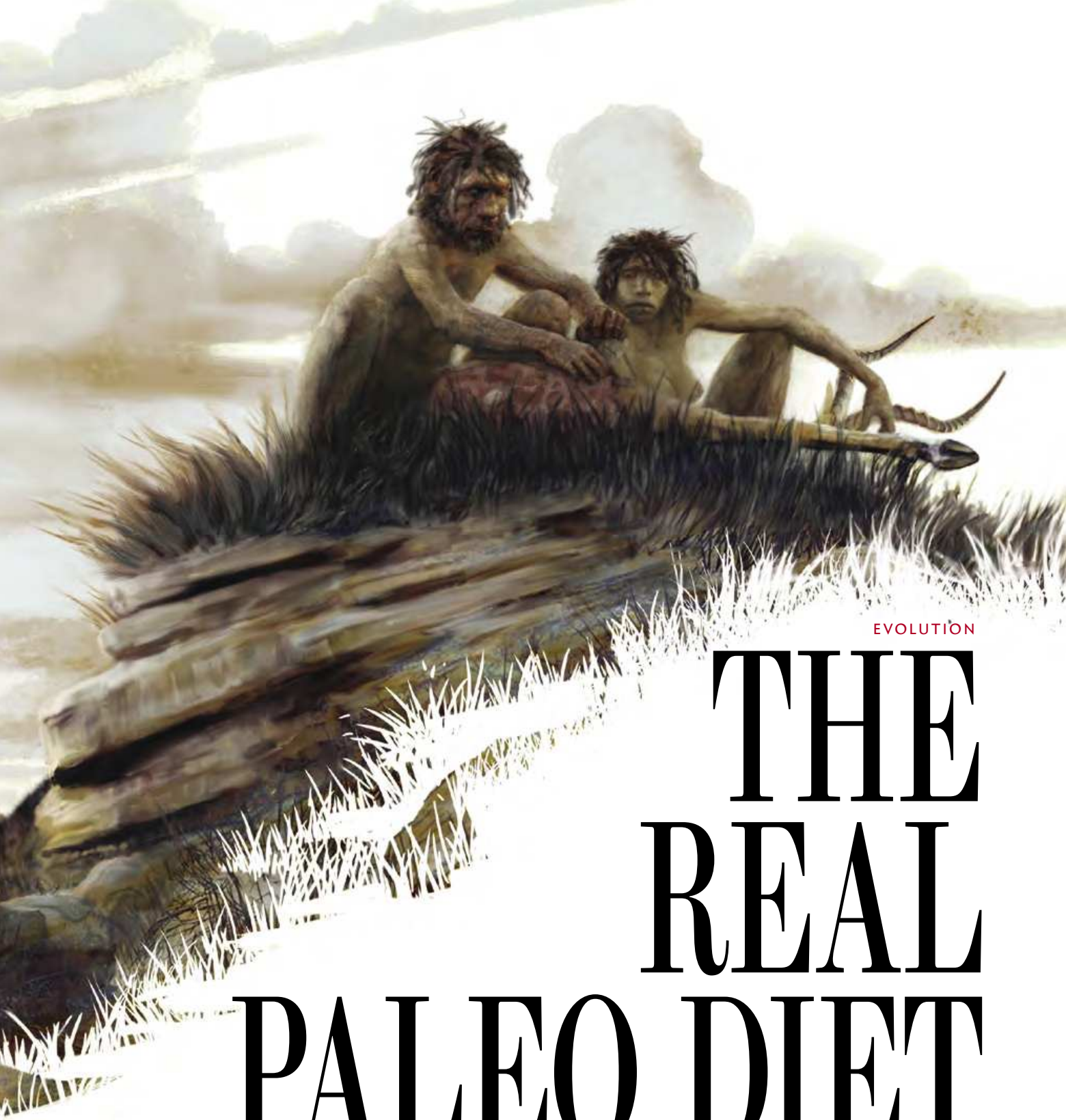
FROM OUR ARCHIVES

The Other Half of the Brain. R. Douglas Fields; April 2004.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American



EVOLUTION

THE REAL PALEO DIET

Microscopic wear patterns on fossil teeth reveal what our ancestors ate—and provide insights into how climate change shaped human evolution

By Peter S. Ungar

Illustration by Jon Foster

© 2018 Scientific American

LATE ONE EVENING IN 1990 AT THE KETAMBE RESEARCH STATION IN INDONESIA'S Gunung Leuser National Park, I sat transcribing notes by the light of a kerosene lamp in my hut on the banks of the Alas River. Something was bothering me. I had come to gather data for my dissertation, documenting what and how the monkeys and apes there ate. The idea was to relate those observations to the sizes, shapes and wear patterns of their teeth. Long-tailed macaques have large incisors and blunt molars—teeth built for eating fruit, according to the received wisdom. But the ones I had been tracking for the past four days seemed to eat nothing but young leaves. I realized then that relations between tooth form and function are more complicated than the textbooks suggest and that the sizes and shapes of an animal's teeth do not dictate what it eats. This might sound like an esoteric revelation, but it has key implications for understanding how animals—including humans—evolved.

IN BRIEF

Paleontologists have traditionally assumed that the sizes and shapes of an animal's teeth dictate its diet. But food availability—which changes seasonally and over larger timescales—is an even bigger factor in what an animal chooses to eat.

Analyses of microscopic use wear and chemical traces of food on fossil teeth are going beyond conventional studies of tooth form to reveal what human ancestors truly ate.

The diet findings, along with paleoenvironmental data, provide fresh insights into how climate change shaped human evolution.

I am a paleontologist, and I earn a living reconstructing the behaviors of extinct species from their fossilized remains. Specifically, I work to discern how animals in the past obtained food from their surroundings and thus how environmental change triggers evolution. That year at Ketambe shaped my way of thinking about primates and the larger community of life that surrounds them. I began to see the biosphere—the part of our planet that harbors life—as a giant buffet of sorts. Animals belly up to the sneeze guard with plates in hand to pick from items available in a given place, at a given time. Each species' place in the forest, and in nature, is defined by the choices it makes.

Teeth play a role in food choice—you need the right utensils. But I learned at Ketambe that availability is even more important. The macaques ate leaves because that is what nature laid out on the biospheric buffet at that time and place. Their diet changed over the course of the year as leaves unfurled, flowers bloomed and fruits ripened with the passing seasons. I began to imagine how changes in food availability over centuries, millennia or longer could affect what a species eats.

Most paleontologists are not used to thinking about life in the past this way. Our field has a long tradition of inferring function from form by assuming that nature selects the best tools for whatever job an organism has to do. If form always followed function, however, macaques would not eat leaves. But how can we detect food choices in the fossil record?

I have spent decades doing exactly that by studying microscopic wear patterns on fossil teeth, including those of a number of human ancestors. Other researchers have analyzed the chemical signatures of food in fossil teeth for dietary clues. These "foodprints," as I call them, reveal the kinds of foods individuals actually ate and have given us a much richer picture of the past than tooth shape alone. Together with insights from the paleoenvironmental record, these findings have allowed us to test some leading hypotheses about the impact of climate change on human evolution. The results refine the classic explanation for how our branch of the human family tree succeeded where others did not.

LIEM'S PARADOX

OBSERVATIONS OF LIVING ANIMALS have revealed numerous creatures that eat foods other than the ones to which they are adapted. While I was at Ketambe, Melissa Remis, now at Purdue University, was gathering diet data on gorillas at Bai Hokou, a lowland rain forest site in the Central African Republic's Dzanga-Ndoki National Park. At that time, most researchers thought that gorillas were dietary specialists that ate stems, leaves and the pith of nonwoody plants such as wild celery. Pioneering gorilla researcher Dian Fossey and others had shown as much in the high-altitude cloud forests of the Virunga Mountains in Uganda and Rwanda. It made sense. Gorillas have very specialized teeth and guts—sharp-crested molars well suited to shearing tough plant parts and a massive hindgut to host microorganisms that help to digest cellulose in fibrous foods. Besides, there was little else to eat at those elevations.

The Virunga mountain gorillas were actually a small, marginal population of just a few hundred indi-

© 2018 Scientific American

viduals living in an extreme habitat, however. What about the 200,000 gorillas living 1,000 miles to the west in the lowland rain forests of the Congo Basin? The gorillas at Bai Hokou told a different story. They seemed to prefer soft, sugary fruits. In fact, Remis saw gorillas walk half a mile or more, right past edible leaves and stems, to get to a fruiting tree. Fibrous foods seemed to dominate their diet only when favored fruits were unavailable. But western lowland gorillas were skittish compared with their cousins in the Virunga Mountains, limiting the amount of data Remis could collect. Some researchers questioned whether gorillas could actually prefer fruits, given their teeth and guts.

There is an old joke: "What do you feed a 400-pound gorilla? Anything it wants." How can we know what a gorilla wants to eat? After Remis returned home from Bai Hokou, she went to the San Francisco Zoo to ask the gorillas themselves. She offered the captive apes a variety of foods, from sweet mango to bitter tamarind, sour lemon and, of course, tough celery. The zoo gorillas clearly preferred sugary, fleshy fruits to tough, fibrous foods, regardless of what their teeth and guts suggested they should eat. This finding confirmed that although gorillas are adapted to the most mechanically and chemically challenging foods they have to eat, these are not their favored foods. Perhaps, then, gorillas in the Virunga Mountains eat tough, fibrous foods year-round not because they prefer them but because they can—and must, given the limited options on the biospheric buffet at such high elevations. Indeed, nearby mountain gorillas that live at lower altitudes prefer to eat fruit when it is available.

A preference for foods other than those to which one is adapted is common enough in the animal kingdom to merit a term for the phenomenon: Liem's paradox. The late Karel Liem of Harvard University observed the paradox first in 1980 in Minckley's cichlid, a freshwater fish endemic to the valley of Cuatro Ciénegas in northern Mexico. One form of this fish has flat, pebblelike teeth in its throat that are seemingly perfectly suited for cracking hard-shelled snails. Yet members of this group swim right past those snails when softer foods are available. Why would an animal evolve teeth specialized for less preferred, rarely eaten items? So long as the hard-object specialization does not preclude consumption of softer foods, it can leave an animal more options when it needs them. The paradox, then, is not so much that individuals avoid the foods to which they are adapted but that specialized anatomy can lead to a more generalized diet.

Other primates exemplify Liem's paradox, including the gray-cheeked mangabey monkeys of Uganda's Kibale National Park. Mangabeys have flat, thickly enameled molars that seem to be specialized for crushing hard, brittle foods. But day after day, month after month, even year after year, Joanna Lambert, now at the University of Colorado Boulder, watched them eat soft, fleshy fruits and young leaves, just like the thinner-toothed red-tailed guenon monkeys that lived alongside them. Then, in the summer of 1997, everything changed. The forest was reeling from an especially severe drought brought on by an El Niño event. Fruits were scarce, leaves were wilting and the monkeys were hungry. The mangabeys ate more bark and hard seeds, but the guenons did not. The mangabeys' specialized teeth and jaws allowed them to fall back on mechanically challenging foods. Even if such adaptations are needed only once or twice in a generation, that can be just what the animals require to get through the lean times.

Specialized anatomy can also relate to preferred foods, though. Sooty mangabeys in Ivory Coast's Taï National Park, for instance, have thick tooth enamel and strong jaws, and they actually prefer hard foods. Much of their foraging time is devoted to scouring the forest floor for seeds of the *Sacoglottis* tree, which have casings that resemble peach pits. Scott McGraw of Ohio State University argues that this practice allows them to avoid competing for food with the 10 other primate species that live alongside them. Just as gorillas vary in

ALAIN HOULE

GRAY-CHEEKED MANGABEYS have flat, thickly enameled molars that appear to be specialized for crushing hard foods. But they fall back on these foods only when the soft fruits and leaves they prefer are unavailable.

© 2018 Scientific American

how often they eat mechanically challenging foods, some mangabeys eat them all the time, and others do so only on rare occasions.

Examples such as these show that primate food choice is complex and depends not just on teeth but also on availability, competition and personal preference. Tooth form can tell us something about what an animal in the past was capable of eating and the most challenging foods its ancestors had to contend with. But for insights into food choices among options that were available on the biospheric buffet, we need foodprints.

Dental microwear, the microscopic scratches and pits that form on a tooth's surface as the result of its use, is a commonly studied type of foodprint. Species that tend to shear or slice tough foods, such as grass-grazing antelopes or meat-eating cheetahs, get long, parallel scratches as opposing teeth slide past one another and abrasives between them are dragged along. Species that crush hard foods, such as nut-eating Taï mangabeys or bone-crunching hyenas, tend to have cratered microwear surfaces, covered in pits of various sizes and shapes.

Because those marks typically wear away and are overwritten in a matter of days, we can learn something about the variety, and perhaps even the proportions, of foods eaten if we consider teeth of individuals sampled at different times and places. The microwear patterns of Kibale mangabeys typically resemble those of soft-fruit eaters, with wispy scratches and fine pits, although a few specimens are more heavily pitted. The teeth of mangabeys from Taï, in contrast, have much more cratered surfaces on average. Despite similar tooth form in the two species, foodprints distinguish them as predicted, based on observations of their diets.

Tooth form can tell us something about what an animal in the past was capable of eating and the most challenging foods its ancestors had to contend with. But for insights into food choices among options that were available on the biospheric buffet, we need foodprints.

ANCIENT MENUS

WITH MICROWEAR PATTERNS from living animals whose dietary habits are known from firsthand observation to guide us, scientists can use microwear on fossil teeth to infer what extinct species ate on a daily basis and gain insight into their food choices. To that end, my colleagues and I have put a lot of effort into analyzing the microwear of human fossils. Our work has generated surprising results.

The human family tree has many branches. Today *Homo sapiens* is the only human species alive, but once upon a time, multiple human species, or hominins, shared the planet. Why our lineage survived when others went extinct is an enduring question. My own foray into this mystery began when I set out to study the diet of members of one of these extinct branches, a group of species belonging to the genus *Paranthropus*. *Paranthropus* lived in eastern and southern Africa between about 2.7 million and 1.2 million years ago, during the Pleistocene epoch. None of its species gave rise to us; rather they were evolutionary experiments that walked alongside our own early ancestors. *Paranthropus* had big, flat, thick-enameled premolars and molars, heavy jaws, and the telltale bony ridges and scars that come from having massive, powerful chewing muscles. These traits are clearly dietary specializations for extreme chewing, so these species seemed to be ideal candidates for microwear analysis. If my collaborators and I could not figure out what they ate, then we had little hope of reconstructing diets of other fossil hominins with less distinctive jaws and teeth.

Paleoanthropologist John Robinson was the first to try to reconstruct the diet of *Paranthropus,* back in 1954. Robinson believed that the large, flat and thickly enameled premolars and molars of *Paranthropus robustus* from South Africa had evolved for grinding plant parts, such as shoots and leaves, berries and tough wild fruits. Chipping on those teeth suggested to him that *P. robustus* ate grit-laden roots and bulbs. The late Phillip Tobias of the University of the Witwatersrand, Johannesburg, saw things differently, arguing in the 1960s that the chips occurred during consumption of hard foods rather than gritty ones. At the time, Tobias was describing a new species of *Paranthropus* from East Africa, *Paranthropus boisei*. On first seeing its skull, he is famously reported to have said, "I have never seen a more remarkable set of nutcrackers."

The idea of a hominin that specialized in nut cracking was born. *Paranthropus* stood in sharp contrast to early *Homo* fossils found in the same sedimentary deposits, with their daintier teeth and jaws, larger brain and emerging stone tool kit for processing food. Researchers came up with a tidy explanation for the differences, dubbed the savanna hypothesis. As grasslands began to spread across Africa, our ancestors came to an evolutionary fork in the road. *Paranthropus* went one way, evolving to specialize on hard, dry savanna plant parts, such as seeds and roots. Early *Homo* went another direction, becoming increasingly versatile, with a more flexible diet that included meat. That dietary flexibility is why we are here today and *Paranthropus* is gone, according to the theory. It was a

© 2018 Scientific American

compelling story, and early microwear studies by Frederick Grine of Stony Brook University in the 1980s showed that the teeth of *P. robustus* do have more microwear pits than those of its own predecessors, seemingly confirming that this cousin of ours specialized in hard, brittle foods.

But in 2005, when my then postdoctoral fellow Rob Scott and I looked again at *P. robustus* microwear using newer technology, another part of the story began to emerge. Yes, *P. robustus* specimens had more pitted, complex microwear surfaces on average, but some of the specimens we studied had less pitted, simpler textures. In fact, microwear in *P. robustus* varied a lot, suggesting that while some ate hard foods in the days before they died, others did not. To put it another way, the specialized anatomy of *P. robustus* did not mean it was a dietary specialist. This was not a new idea. David Strait, now at Washington University in St. Louis, and Bernard Wood of George Washington University had the year before speculated that *Paranthropus* may well have been an ecological generalist with a flexible diet, based largely on indirect evidence. But our work provided direct evidence for Liem's paradox among the hominins.

A bigger surprise came in 2008, when my colleagues and I looked at the microwear textures of *P. boisei*. This was Tobias's nutcracker, the species with the largest teeth, heaviest jaws and thickest enamel of all the hominins. I expected *P. boisei*'s teeth to have microwear akin to that of the sooty mangabey's, cratered like the surface of the moon. They did not. Surface after surface had wispy scratches running every which way. Not only were these critters not hard-object specialists, but their microwear showed no sign at all of hard foods. The nutcracker hypothesis seemed to fall like a house of cards in a stiff wind. So what was *P. boisei* eating with those big, flat teeth? That would have to wait on another set of foodprints: carbon isotope ratios.

Distinctive chemical signatures of foods that provide the raw materials used to build the body are sometimes preserved in teeth. Like microwear, these chemical clues can be read and decoded. For example, compared with trees and bushes, tropical grasses have a higher proportion of carbon atoms with seven neutrons rather than the usual six; the teeth of animals that eat tropical grasses have predictably more "heavy" carbon as a result.

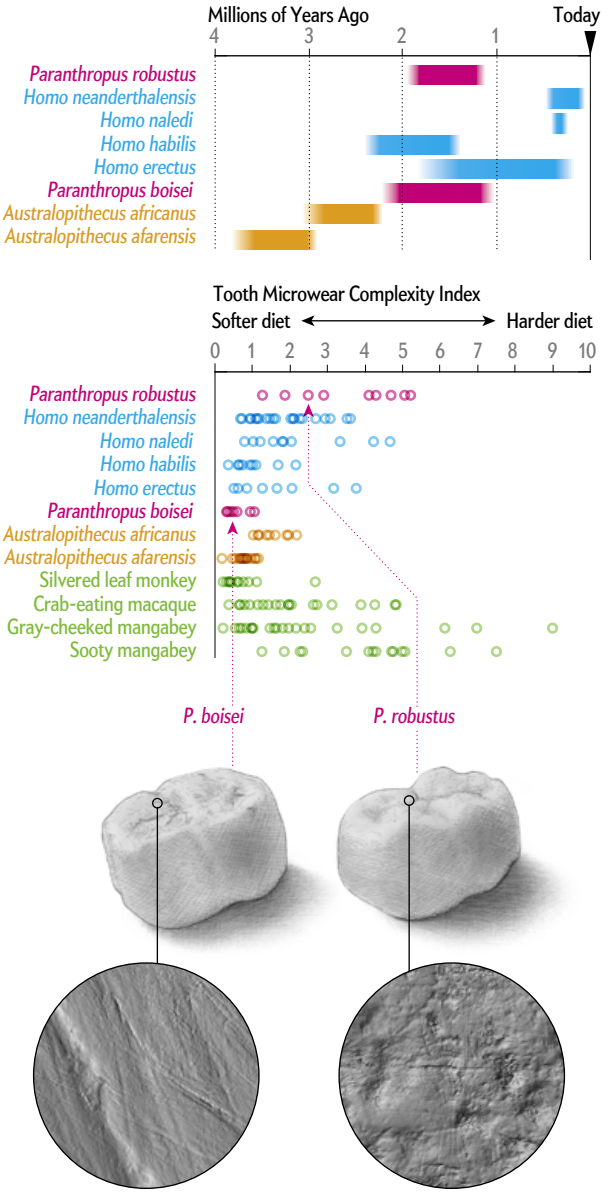
Carbon isotope ratios of *P. robustus* teeth indicate a diet dominated by tree and bush products but with a hearty helping of tropical grasses or sedges. This finding is consistent with a broad-based diet. But *P. boisei* shows a very different pattern, with carbon isotope ratios suggesting that grasses or sedges made up at least three quarters of its diet.

This result came as a surprise to many paleoanthropologists. A cowlike hominin? Surely no self-respecting member of our family tree would earn its living eating grass! But it made sense to me. These species debuted just as grasslands were spreading across eastern and southern Africa, and the biospheric buffet table was becoming covered in turf. If *P. boisei* was grinding grass or sedge products with its big, flat teeth and powerful jaws rather than crushing hard, brittle foods, that should leave exactly the microwear texture pattern my colleagues and I found. Such a diet would also explain why *P. boisei* wore down its molars so quickly.

PETER S. UNGAR (*data and microwear simulations*)

Foodprints

Microscopic scratches and pits form on teeth as a result of their use. Studies of these microwear patterns in living animals show that species that chew soft and tough foods such as grass, for example, get long, parallel scratches on their teeth; those that crush hard and brittle foods such as nuts get pits. Paleontologists have inferred the diet of extinct human species, including *Paranthropus robustus* and *Paranthropus boisei,* based on the microwear textures on fossil teeth.



Previous studies based on tooth form concluded that *P. robustus* ate tough plants and that *P. boisei* specialized in cracking nuts. Microwear analysis, however, reveals that *P. robustus* had a complex pattern of pits and scratches indicative of a dietary generalist. *P. boisei,* for its part, had none of the pits one would expect to see in a nutcracker. Subsequent chemical analyses indicated that *P. boisei* ate mostly grasses or sedges.

Illustration by Portia Sloan Rollings (teeth), Graphics by Jen Christiansen

© 2018 Scientific American

JOHN R. FOSTER *Science Source*

You would never know it by just looking at the shapes of their huge, flat teeth, but foodprints suggest that the two *Paranthropus* species used their specialized anatomy in different and unexpected ways. Like the Kibale mangabeys, *P. robustus* seems to have had a generalized diet that included some hard objects. But for *P. boisei,* the relation between teeth and diet seems to have been very different from anything we see in primates today. Big, flat teeth are far from ideal for shredding grass, but one works with what one has. And so long as a grinding platform is better than what hominins had before, it would be selected for even if it is not optimal for the task at hand.

Microwear of our direct ancestors—those in the *Homo* genus—points to a decidedly different dietary strategy. My colleagues and I have looked at two early species: the more "primitive" *Homo habilis,* a smaller-brained hominin that retained some features related to life in the trees, and *Homo erectus,* a larger-brained hominin committed to the ground. Our samples are small because microwear requires pristine teeth, and there are just not that many of them. But they show an interesting pattern. Compared with *Australopithecus afarensis,* its putative ancestor, and *P. boisei,* which lived alongside it, *H. habilis* has a somewhat broader range of microwear textures, from complex pitted surfaces to simple scratched ones. The finding hints that *H. habilis* ate a wider range of foods than either its predecessors or its contemporaries. Its successor *H. erectus* has even more variable microwear textures, perhaps suggesting a broader diet still.

These results fit neatly with a leading model of how climate change shaped human evolution that has superseded the savanna hypothesis. Work on climate data from deep ocean cores in the mid-1990s by the late geologist Nicholas Shackleton showed there was more to the story of climate change than the savanna hypothesis supposed. Conditions did become cooler and drier over the long term, but there were also short-term climate swings, and those swings became more and more intense over the course of human evolution.

DIVERGING DIETS: **Whereas *Paranthropus boisei* (*1*) specialized in eating grasses or sedges, its contemporary *Homo habilis* (*2*) appears to have had a broader diet.**

Rick Potts of the Smithsonian Institution reasoned that this unstable climate pattern should favor more versatile species, including hominins—an idea that became known as the variability selection hypothesis. Pleistocene Africa was no place to be a picky eater. For Potts, it was not so much the spread of savanna grasses but the need for flexibility that drove human evolution. In this light, *Homo*'s larger brain and stone tools for processing a variety of foods make sense. They would have allowed our ancestors to survive increasingly intense environmental swings and to keep up as nature more quickly swapped items on and off the biospheric buffet. The increasing variation in microwear complexity from *A. afarensis* to *H. habilis* to *H. erectus* just might be direct evidence of variability selection.

Potts's idea has held up pretty well in the two decades since he first presented it, although others have built on it, and new details have emerged about how changes to Earth's landscapes and in its orbit around the sun have combined to create the conditions under which humans evolved. For example, in 2009 Mark Maslin of University College London and Martin Trauth of the University of Potsdam in Germany suggested that climate swings filled and emptied the

© 2018 Scientific American

spreading lakes in eastern Africa, disrupting life in the rift basins. This flux may have led to fragmentation and dispersal of hominin populations, fueling human evolution. The ability to pursue a more variable diet would have aided survival in such turbulent times.

APPETITE AND EVOLUTION

ALTHOUGH THE AVAILABLE EVIDENCE allows scientists to paint a plausible picture of how early hominins adapted to their changing world, we can only do so with the broadest of brushstrokes. The biggest challenge to understanding how climate change drives evolution is matching specific climate events in the past to changes in the fossil record.

Local environments react to global and even regional climate change in different ways, and our fossil record is simply not complete enough to tell exactly where and when particular species appeared and disappeared. We can be off by 1,000 miles and 100,000 years or more. We might be able to tie the extinction or evolution of a given species to a massive, catastrophic event in Earth's history, such as the asteroid impact in the Yucatán Peninsula that killed off the dinosaurs 66 million years ago. But the climate-related events we associate with human evolution are very different—repeated cycles of cool-dry conditions followed by warm-wet ones. The fact that hominins were probably flexible species capable of adjusting to a broad range of habitats and the foods available within them further obscures the picture. Our best shot at understanding how hominins responded to changing environments thus lies in the more recent past, in places that are exceptionally well studied.

Research published by Sireen El Zaatari of the University of Tübingen in Germany, Kristin Krueger of Loyola University Chicago and their colleagues over the past two years shows how this approach might work. Their studies of the microwear of Neandertals and the anatomically modern humans that supplanted them in Eurasia allow us to revisit the long-standing mystery of this replacement from a fresh perspective. Neandertals ruled Europe and western Asia between about 400,000 and 40,000 years ago. Then they were gone. Paleoanthropologists have been debating what happened and why for more than a century, and even today there is little consensus.

Although popular science often tells a tale of brutish Neandertals living in near-glacial conditions, swaddled in animal hides and gorging lustfully on mammoth and woolly rhinoceros meat, it was not always like that. Neandertals inhabited a wide range of habitats, from cold, dry steppes to warmer, wetter woodlands, and conditions varied over time and space. Recent studies of their molars show that Neandertals living in more wooded or mixed settings had complex pitted microwear, suggesting that they ate more hard, brittle and perhaps abrasive plant foods. Neandertals that dwelled on the open steppes, in contrast, have less complex molar microwear, which El Zaatari and her colleagues argue reflects a less variable diet composed primarily of soft meat. Krueger, for her part, found differences in incisor microwear between the two groups; she thinks the differences stem from the steppe Neandertals having used their incisors to aid in processing animal hides and the forest Neandertals having eaten a greater variety of foods. Intriguingly these differences hold whether one considers earlier Neandertals or later ones. It seems that Neandertals were flexible feeders with diets that tracked to habitat and associated food availabilities.

The pattern is different, though, for anatomically modern people living in Europe during the last ice age. There is not much difference in molar microwear between those from open habitats and those who occupied habitats containing a mix of open and wooded vegetation, whether one considers earlier or later individuals. Perhaps early modern humans were better able to acquire their preferred foods than Neandertals were when faced with environmental change.

FOOD FOR THOUGHT

STUDIES OF EARLY HUMAN DIETS bear on what people today should eat to be healthy—though perhaps not in the manner popularly envisioned. "Paleolithic diet" gurus argue that we should eat the kinds of foods our ancestors evolved to eat. Many chronic degenerative diseases have been linked to a mismatch between our diets and the fuels our bodies were "designed" to burn, they contend. And it certainly cannot hurt to remind ourselves every now and again that our distant forebears did not eat corn dogs or milkshakes.

That does not mean that we should look to follow a specific Paleolithic diet, however. Foodprints teach us that early hominin diets varied over time and space and that we mostly likely evolved to be flexible eaters, driven by ever changing climates, habitats and food availability. In other words, there was no single ancestral human diet for us to replicate. Dietary versatility allowed our ancestors to spread across the planet and find something to eat on all of Earth's myriad biospheric buffets. It was the key to our evolutionary success. SA

Peter S. Ungar is a paleontologist at the University of Arkansas. His research focuses on diet and feeding adaptations in living and fossil primates, including human ancestors. His most recent book is *Evolution's Bite* (Princeton University Press, 2017).

MORE TO EXPLORE

Dental Microwear and Diet of the Plio-Pleistocene Hominin *Paranthropus boisei*. Peter S. Ungar et al. in *PLOS ONE*, Vol. 3, No. 4, Article No. e2044; April 30, 2008.

The Diets of Early Hominins. Peter S. Ungar and Matt Sponheimer in *Science*, Vol. 334, pages 190–193; October 14, 2011.

Neandertal versus Modern Human Dietary Responses to Climatic Fluctuations. Sireen El Zaatari et al. in *PLOS ONE*, Vol. 11, No. 4, Article No. e0153277; April 27, 2016.

FROM OUR ARCHIVES

Climate Shocks. Peter B. deMenocal; September 2014.

scientificamerican.com/magazine/sa

© 2018 Scientific American

VIEW OF MARS'S SOUTH POLAR ICE CAP and the surrounding area was captured by the European Space Agency's Mars Express spacecraft on February 25, 2015. Recent observations from Mars Express have revealed a lake of liquid water lurking beneath the ice cap.

© 2018 Scientific American

SPACE

Watery Red Planet?

Radar observations have revealed what appears to be a buried lake on Mars, the first ever stable reservoir of liquid water found on the planet

By Lee Billings

© 2018 Scientific American

LOCATED AT THE EDGE OF A MORE THAN THREE-BILLION-YEAR-OLD ICE CAP covering Mars's south pole, the region known as Planum Australe would rank high on any list of the Red Planet's least interesting locales. Frozen, flat and featureless, it seemingly offers little more than windblown dust and drifts of crystallized carbon dioxide for any aspiring explorer to see. Unless, that is, one could somehow peer deep underneath its frigid surface to the base of the ice cap some 1.5 kilometers below, where a lake of liquid water nearly three times larger than the island of Manhattan may lurk.

Discovered by a team of Italian scientists using three years' worth of data from the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) instrument on the European Space Agency's Mars Express orbiter, the potential lake is at least a few meters deep and might be a fixed, steady feature of the subsurface. If confirmed, this would be the first known reservoir of liquid water on present-day Mars—a keystone in the search for past or even present life on the Red Planet, potentially offering fresh clues about how Earth's neighbor so profoundly transformed billions of years ago from a warmer, wetter world to its current freeze-dried state. Announced July 25 at a press conference in Rome, the results are detailed in a study in the July 26 edition of *Science*. Although this is just one detection, the team wrote, "there is no reason to conclude that the presence of subsurface water on Mars is limited to a single location."

"The presence of a body of liquid water beneath Mars's south polar cap has various implications, opening new possibilities for the existence of microorganisms in the Martian environment," says Sebastian Lauro, a study co-author based at Roma Tre University in Rome. "Moreover, it provides a valuable confirmation that the water that once flowed abundantly over the Martian surface in the form of seas, lakes and rivers filled the voids in the subsurface."

For the past 12 years MARSIS has mapped the Martian underground using beams of low-frequency radar pulses, which can penetrate up to several kilometers below the surface. Although they pass relatively unscathed through most substances, these pulses reflect back up to the spacecraft each time they encounter boundaries between different materials, such as the interface of ice and bedrock. That reflection is particularly strong at interfaces with liquid water and shows up as a distinctively bright spot in visualizations of the data. Following up on preliminary detections of bright spots beneath Mars's southern ice cap dating back to 2007, the Italian team reprogrammed MARSIS to employ a more intensive scanning mode, then surveyed Planum Australe 29 times with the instrument between 2012 and 2015. Time and time again across the entire observing campaign the new MARSIS readings revealed a consistent 20-kilometer-wide bright spot nestled in a bowl-like depression beneath the ice cap in Planum Australe—a feature consistent with a sizable body of liquid water (or, to be fair, with water-saturated sediments more akin to subterranean sludge). The team then spent almost a year analyzing the data and another two years writing their paper and attempting to rule out nonaqueous explanations for what they had seen.

Billions of years ago Mars was a much more Earth-like place, where water pooled in seas, carved enormous canyons and bubbled from hot springs. Life, many astrobiologists speculate, may have had no difficulty getting started there. But early in its promising existence the planet somehow lost its way, transforming into a desiccated orb of dried-up oceans, rivers and lakes. Robotic missions to the planet's surface still find surprising echoes of that bygone time, such as patches of water-ice frost forming on rocks as well as water droplets condensing like dew on a lander's leg. Orbiters, too, have glimpsed what might be rivulets of water flowing down sun-bathed crater walls at the height of Martian summer. Perhaps life has also managed to endure in some diminished, limited way. But, if so, it would have to contend with a world in which all moisture quickly vanishes in the thin, cold air, leaving the surface dry as a bone. Still, the water that once flowed across the land had to go somewhere. Some of it was likely lost to space, due to Mars's diminutive gravitational field, but a significant fraction of the planet's aqueous inventory never really left, instead just freezing belowground. Now it appears not all of that buried watery wealth is frozen after all.

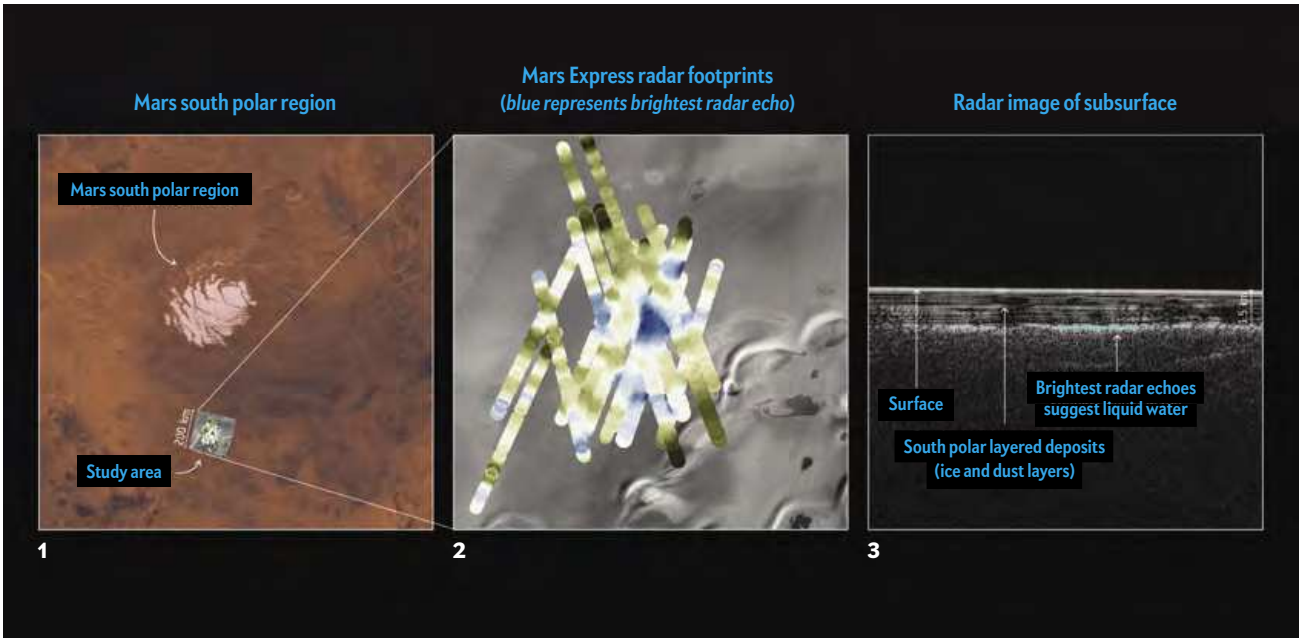
"The really exciting thing is that this is a stable body of liquid water that was observed in the radar data over three years, not just droplets that have been observed over a short period of time," says Anja Diez, a glaciologist at the Norwegian Polar Institute, who wrote an accompanying commentary about the discovery. The subsurface lake, Diez says, may be similar to those found via radar-sounding on Earth underneath ice sheets in Antarctica and Greenland.

Whether below an Earthly glacier or a Martian ice cap, the mechanism for melting is much the same: heat trickling up combines with the immense bulk of an insulating blanket of material pressing down to form lakes of meltwater. On Earth those lakes are often connected by channels, forming branching riverlike networks of water that extend across vast spaces below the ice. In the late 1980s Steve Clifford, a researcher now at the Planetary Science Institute, began exploring how similar hydrological activity could occur under both Mars's southern and northern polar caps and whether it might feed meltwater into worldwide aquifers he hypothesized should exist beneath the planet's permafrost. Clifford's models suggest huge amounts of liquid water could still be hidden in the planet's depths, providing a globe-spanning refuge for any life that retreated from the ever more inhospitable surface long ago.

"This finding is potentially of enormous significance," says Clifford, who was not involved with the study. "Based on analogy with Earth, if water still exists in the subsurface, there is no reason to believe that life which arose on Mars and evolved for underground conditions could not persist there into the present day.… If you do have liquid water as shallow as 1.5 kilometers beneath the surface [at Planum Australe], then liquid water

PRECEDING PAGES: ESA, DLR AND FREE UNIVERSITY BERLIN (CC BY-SA 3.0 IGO)

© 2018 Scientific American



RADAR STUDIES of Mars's Planum Australe (*1*) have revealed evidence of a lake (*large blue spot in 2*) buried under 1.5 kilometers of ice and dirt (*3*).

is also likely to be present at greater depths here. And if you have conditions for life in one area of the planet that is in hydraulic continuity with other areas where liquid water also exists, you could have a very substantial subsurface biosphere that has survived since the planet's early history."

That life, however, would have to contend with another key factor making its aquatic environs possible: mineral salts that leach out of rocks and sediments to act as antifreeze. Suffusing the meltwater, the salts would create brines that remain liquid far below the typical freezing point of pure water. Such salts are known to exist in abundance in some Martian rocks and are the most likely cause of the dewlike droplets and crater-wall rivulets previously observed on the planet's surface. But Clifford holds out hope subsurface geothermal hotspots like those that power volcanoes and hot springs on Earth could sufficiently heat portions of the Martian underworld to allow liquid reservoirs to exist there without the need for life-sabotaging salt levels. Such a hot spot could, in fact, be responsible for the MARSIS team's newfound lake.

These conjectures must, for now, remain untested. The MARSIS instrument lacks sufficient sensitivity and resolution to clearly determine the thickness of this deposit or whether it is in fact connected to other similar bodies, although studies using more advanced radar instruments on an as yet unbuilt next generation of orbiters could clarify such details. For that matter, the detection itself is as yet uncorroborated: Another radar-sounding instrument called SHARAD (for "shallow radar") onboard NASA's Mars Reconnaissance Orbiter has also repeatedly scanned Planum Australe and other regions in search of buried water. But its beams cannot penetrate as deeply, and it has not replicated MARSIS's detection. Even so, according to Bruce Campbell, a SHARAD team member and senior scientist at the Smithsonian Institution, "it is possible that a targeted effort [by SHARAD] to collect data on more tracks across this region could build up enough echo strength to detect the reflector seen by MARSIS."

Ultimately "ground truth" may be required to solve the mystery of just how much water is locked away in Mars's underworld. But there are at present no public or private plans to build or launch missions, robotic or human, that would be capable of drilling or melting down to the depths probably required to directly sample meltwater anywhere on the planet—which, Clifford says, might be for the best.

"The possibility that life could currently exist somewhere beneath the polar ice reinforces the point that we must take care that our investigations do not needlessly contaminate Mars," he says. "That could not only make the result of any life-detection experiment ambiguous but could also contaminate a habitat that may be interconnected on a global basis, leading to a very serious impact on any native biosphere. I worry that unless we are very careful, we could end up responsible for the extinction of the very first life we ever detect on another planet." SA

Lee Billings is an associate editor for *Scientific American*. He covers space and physics.

MORE TO EXPLORE

Subsurface Radar Sounding of the South Polar Layered Deposits of Mars. Jeffrey J. Plaut in *Science*, Vol. 316, pages 92–95; April 6, 2007.

Mars Express Detects Water Buried under the South Pole of Mars. Press release from the European Space Agency, released July 25, 2018, 4:00 P.M. www.esa.int/spaceinimages/Images/2018/07/

Radar Evidence of Subglacial Liquid Water on Mars. Roberto Orosei et al. in *Science*, Vol. 361, pages 490–493; August 3, 2018.

FROM OUR ARCHIVES

How to Search for Life on Mars. Christopher P. McKay and Victor Parro García; September 2017.

scientificamerican.com/magazine/sa

CONTEXT MAP: NASA AND VIKING; THEMIS BACKGROUND: NASA, JPL-CALTECH AND ARIZONA STATE UNIVERSITY; MARSIS DATA: ESA, NASA, JPL, ASI AND UNIVERSITY OF ROME; R. OROSEI ET AL. 2018

© 2018 Scientific American

© 2018 Scientific American

MATHEMATICS

The Un(solv)able Problem

After a years-long intellectual journey, three mathematicians have discovered that a problem of central importance in physics is impossible to solve—and that means other big questions may be undecidable, too

By Toby S. Cubitt, David Pérez-García and Michael Wolf

Illustration by Mark Ross Studios

© 2018 Scientific American

THE THREE OF US WERE SITTING TOGETHER IN A CAFÉ IN SEEFELD, A SMALL TOWN DEEP in the Austrian Alps. It was the summer of 2012, and we were stuck. Not stuck in the café—the sun was shining, the snow on the Alps was glistening, and the beautiful surroundings were sorely tempting us to abandon the mathematical problem we were stuck on and head outdoors. We were trying to explore the connections between 20th-century mathematical results by Kurt Gödel and Alan Turing and quantum physics. That, at least, was the dream. A dream that had begun back in 2010, during a semester-long program on quantum information at the Mittag-Leffler Institute near Stockholm.

Some of the questions we were looking into had been explored before by others, but to us this line of research was entirely new, so we were starting with something simple. Just then, we were trying to prove a small and not very significant result to get a feel for things. For months now, we had a proof (of sorts) of this result. But to make the proof work, we had to set up the problem in an artificial and unsatisfying way. It felt like changing the question to suit the answer, and we were not very happy with it. Picking the problem up again during the break after the first session of talks at the workshop in Seefeld that had brought us together in 2012, we still could not see any way around our problems. Half-jokingly, one of us (Michael Wolf) asked, "Why don't we prove the undecidability of something people really care about, like the spectral gap?"

At the time, we were interested in whether certain problems in physics are "decidable" or "undecidable"—that is, can they ever be solved? We had gotten stuck trying to probe the decidability of a much more minor question, one few people care about. The "spectral gap" problem Michael was proposing that we tackle (which we will explain later) was one of central importance to physics. We did not know at the time whether this problem was or was not decidable (although we had a hunch it was not) or whether we would be able to prove it either way. But if we could, the results would be of real relevance to physics, not to mention a substantial mathematical achievement. Michael's ambitious suggestion, tossed off almost as a jest, launched us on a grand adventure. Three years and 146 pages of mathematics later, our proof of the undecidability of the spectral gap was published in *Nature*.

To understand what this means, we need to go back to the beginning of the 20th century and trace some of the threads that gave rise to modern physics, mathematics and computer science. These disparate ideas all lead back to German mathematician David Hilbert, often regarded as the greatest figure of the past 100 years in the field. (Of course, no one outside of mathematics has heard of him. The discipline is not a good route to fame and celebrity, although it has its own rewards.)

THE MATHEMATICS OF QUANTUM MECHANICS

HILBERT'S INFLUENCE on mathematics was immense. Early on, he developed a branch of mathematics called functional analysis—in particular, an area known as spectral theory, which would end up being key to the question within our proof. Hilbert was interested in this area for purely abstract reasons. But as so often hap-

IN BRIEF

Kurt Gödel famously discovered in the 1930s that some statements are impossible to prove true or false—they will always be "undecidable."

Mathematicians recently set out to discover whether a certain fundamental problem in quantum physics—the so-called spectral gap question—falls into this category. The spectral gap refers to the energy difference between the lowest energy state a material can occupy and the next state up.

After three years of blackboard brainstorming, midnight calculating and much theorizing over coffee, the mathematicians produced a 146-page proof that the spectral gap problem is, in fact, undecidable. The result raises the possibility that other important questions may likewise be unanswerable.

© 2018 Scientific American

pens, his mathematics turned out to be exactly what was necessary to understand a question that was perplexing physicists at the time.

If you heat a substance up, it begins to glow as the atoms in it emit light (hence the phrase "red hot"). The yellow-orange light from sodium street lamps is a good example: sodium atoms predominantly emit light at a wavelength of 590 nanometers, in the yellow part of the visible spectrum. Atoms absorb or release light when electrons within them "jump" between energy levels, and the precise frequency of that light depends on the energy gap between the levels. The frequencies of light emitted by heated materials thus give us a "map" of the gaps between the atom's different energy levels. Explaining these atomic emissions was one of the problems physicists were wrestling with in the first half of the 20th century. The question led directly to the development of quantum mechanics, and the mathematics of Hilbert's spectral theory played a prime role.

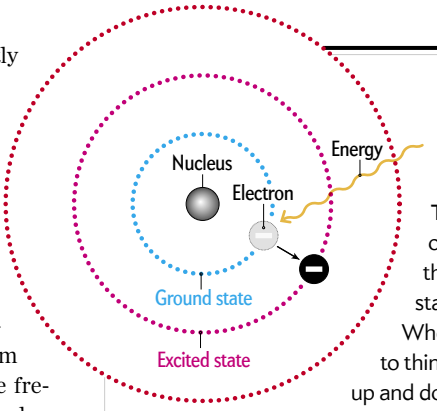
One of these gaps between quantum energy levels is especially important. The lowest possible energy level of a material is called its ground state. This is the level it will sit in when it has no heat. To get a material into its ground state, scientists must cool it down to extremely low temperatures in a laboratory. Then, if the material is to do anything other than sit in its ground state, something must excite it to a higher energy. The easiest way is for it to absorb the smallest amount of energy it can, just enough to take it to the next energy level above the ground state—the first excited state. The energy gap between the ground state and this first excited state is so critical that it is often just called the spectral gap.

In some materials, there is a large gap between the ground state and the first excited state. In other materials, the energy levels extend all the way down to the ground state without any gaps at all. Whether a material is "gapped" or "gapless" has profound consequences for its behavior at low temperatures. It plays a particularly significant role in quantum phase transitions.

A phase transition happens when a material undergoes a sudden and dramatic change in its properties. We are all very familiar with some phase transitions—such as water transforming from its solid form of ice into its liquid form when heated up. But there are more exotic quantum phase transitions that happen even when the temperature is kept extremely low. For example, changing the magnetic field around a material or the pressure it is subjected to can cause an insulator to become a superconductor or cause a solid to become a superfluid.

How can a material go through a phase transition at a temperature of absolute zero (−273.15 degrees Celsius), at which there is no heat at all to provide energy? It comes down to the spectral gap. When the spectral gap disappears—when a material is gapless—the energy needed to reach an excited state becomes zero. The tiniest amount of energy will be enough to push the material through a phase transition. In fact, thanks to the weird quantum effects that dominate physics at these very low temperatures, the material can temporarily "borrow" this energy from nowhere, go through a phase transition and "give" the energy back. Therefore, to understand quantum phase transitions and quantum phases, we need to determine when materials are gapped and when they are gapless.

Because this spectral gap problem is so fundamental to understanding quantum phases of matter, it crops up all over the place in theoretical physics. Many famous and long-standing

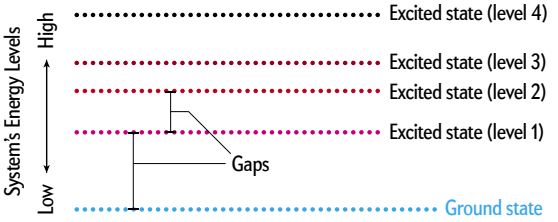


The Spectral Gap

The authors' mathematical proof took on the question of the "spectral gap"—the jump in energy between the ground state and first excited state of a material. When we think of energy states, we tend to think of electrons in atoms, which can jump up and down between energy levels. Whereas in atoms there is always a gap between such levels, in larger materials made of many atoms, there is sometimes no distance between the ground state and the first excited state: even the smallest possible amount of energy will be enough to push the material up an energy level. Such materials are called "gapless." The authors proved that it will never be possible to determine whether all materials are gapped or gapless.

Gapped System

There are discrete gaps between each energy level, and the material must reach a certain energy to make the leap to the next level.



Gapless System

No expanse separates the ground state and first excited state, and the material may become excited with just the tiniest input of energy.

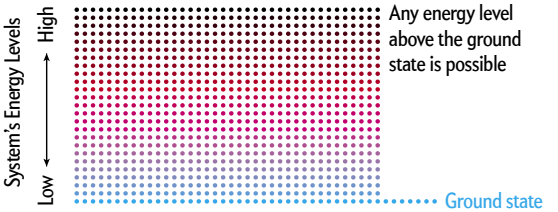


Illustration by Jen Christiansen

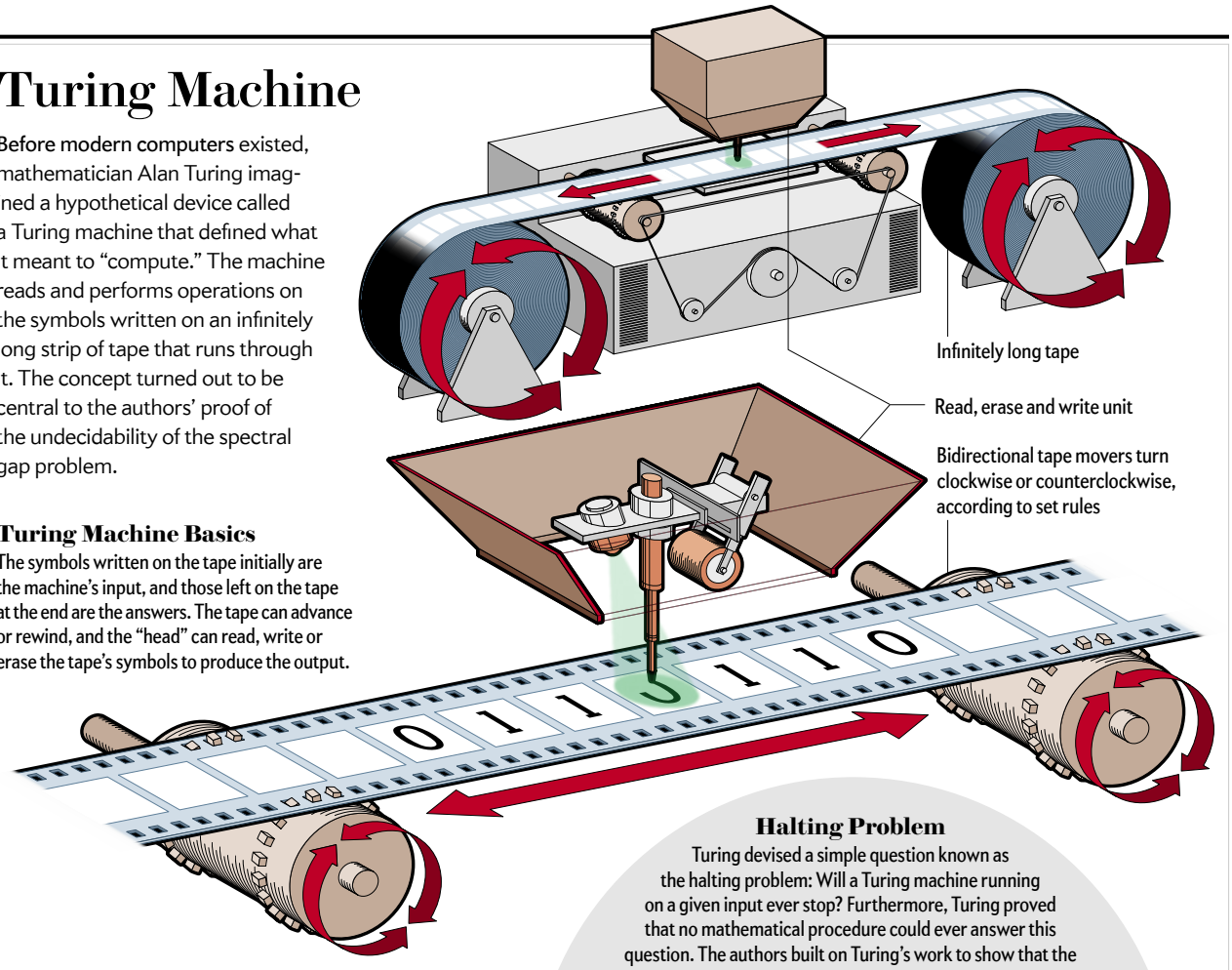
© 2018 Scientific American

Turing Machine

Before modern computers existed, mathematician Alan Turing imagined a hypothetical device called a Turing machine that defined what it meant to "compute." The machine reads and performs operations on the symbols written on an infinitely long strip of tape that runs through it. The concept turned out to be central to the authors' proof of the undecidability of the spectral gap problem.

Turing Machine Basics

The symbols written on the tape initially are the machine's input, and those left on the tape at the end are the answers. The tape can advance or rewind, and the "head" can read, write or erase the tape's symbols to produce the output.


Infinitely long tape
Read, erase and write unit
Bidirectional tape movers turn clockwise or counterclockwise, according to set rules
0 1 1 1 1 0

Halting Problem

Turing devised a simple question known as the halting problem: Will a Turing machine running on a given input ever stop? Furthermore, Turing proved that no mathematical procedure could ever answer this question. The authors built on Turing's work to show that the spectral gap is similar to the halting problem and is likewise undecidable.

open problems in condensed matter physics boil down to solving this problem for a specific material. A closely related question even crops up in particle physics: there is very good evidence that the fundamental equations describing quarks and their interactions have a "mass gap." Experimental data from particle colliders such as the Large Hadron Collider near Geneva support this notion, as do massive number-crunching results from supercomputers. But proving the idea rigorously from the theory seems to be extremely difficult. So difficult, in fact, that this problem, called the Yang-Mills mass gap problem, has been named one of seven Millennium Prize problems by the Clay Mathematics Institute, and anyone who solves it is entitled to a $1-million prize. All these problems are particular cases of the general spectral gap question. We have bad news for anyone trying to solve them, though. Our proof shows that the general problem is even trickier than we thought. The reason comes down to a question called the *Entscheidungsproblem*.

UNANSWERABLE QUESTIONS

BY THE 1920S Hilbert had become concerned with putting the foundations of mathematics on a firm, rigorous footing—an endeavor that became known as Hilbert's program. He believed that whatever mathematical conjecture one might make, it will in principle be possible to prove either that it is true or that it is false. (It had better not be possible to prove that it is both, or something has gone very wrong with mathematics!) This idea might seem obvious, but mathematics is about establishing concepts with absolute certainty. Hilbert wanted a rigorous proof.

In 1928 he formulated the *Entscheidungsproblem*. Although it sounds like the German sound for a sneeze, in English it translates to "the decision problem." It asks whether there is a procedure, or "algorithm," that can decide whether mathematical statements are true or false.

For example, the statement "Multiplying any whole number by 2 gives an even number" can easily be proved true, using basic logic and arithmetic. Other statements are less clear. What about the following example? "If you take any whole number, and repeatedly multiply it by 3, and add 1 if it's odd or divide it by 2 if it's even, you always eventually reach the number 1." (Have a think about it.)

Unfortunately for Hilbert, his hopes were to be dashed. In 1931 Gödel published some remarkable results now known as his incompleteness theorems. Gödel showed that there are perfectly reasonable mathematical statements about whole numbers that can be neither proved nor disproved. In a sense, these statements are beyond the reach of logic and arithmetic. And

Illustration by Ben Gilliland

© 2018 Scientific American

he proved this assertion. If that is hard to wrap your head around, you are in good company. Gödel's incompleteness theorems shook the foundations of mathematics to the core.

Here is a flavor of Gödel's idea: If someone tells you, "This sentence is a lie," is that person telling the truth or lying? If he or she is telling the truth, then the statement must indeed be a lie. But if he or she is lying, then it is true. This quandary is known as the liar paradox. Even though it appears to be a perfectly reasonable English sentence, there is no way to determine whether it is true or false. What Gödel managed to do was to construct a rigorous mathematical version of the liar paradox using only basic arithmetic.

The next major player in the story of the *Entscheidungsproblem* is Alan Turing, the English computer scientist. Turing is most famous among the general public for his role in breaking the German Enigma code during World War II. But among scientists, he is best known for his 1937 paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*." Strongly influenced by Gödel's result, the young Turing had given a negative answer to Hilbert's *Entscheidungsproblem* by proving that no general algorithm to decide whether mathematical statements are true or false can exist. (American mathematician Alonzo Church also independently proved this just before Turing. But Turing's proof was ultimately more significant. Often in mathematics, the proof of a result turns out to be more important than the result itself.)

To solve the *Entscheidungsproblem,* Turing had to pin down precisely what it meant to "compute" something. Nowadays we think of computers as electronic devices that sit on our desk, on our lap or even in our pocket. But computers as we know them did not exist in 1936. In fact, "computer" originally meant a person who carried out calculations with pen and paper. Nevertheless, computing with pen and paper as you did in high school is mathematically no different to computing with a modern desktop computer—just much slower and far more prone to mistakes.

Turing came up with an idealized, imaginary computer called a Turing machine. This very simple imaginary machine does not look like a modern computer, but it can compute everything that the most powerful modern computer can. In fact, any question that can ever be computed (even on quantum computers or computers from the 31st century that have yet to be invented) could also be computed on a Turing machine. It would just take the Turing machine much longer.

A Turing machine has an infinitely long ribbon of tape and a "head" that can read and write one symbol at a time on the tape, then move one step to the right or left along it. The input to the computation is whatever symbols are originally written on the tape, and the output is whatever is left written on it when the Turing machine finally stops running (halts). The invention of the Turing machine was more important even than the solution to the *Entscheidungsproblem*. By giving a precise, mathematically rigorous formulation of what it meant to make a computation, Turing founded the modern field of computer science.

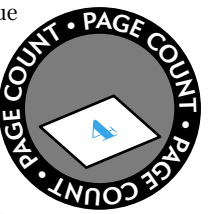
Having constructed his imaginary mathematical model of a computer, Turing then went on to prove that there is a simple question about Turing machines that no mathematical procedure can ever decide: Will a Turing machine running on a given input ever halt? This question is known as the halting problem. At the time, this result was shocking. Mathematicians have become accustomed to the fact that any conjecture we are working on could be provable, disprovable or undecidable.

WHERE WE COME IN

IN OUR RESULT, we had to tie all these disparate threads back together. We wanted to unite the quantum mechanics of the spectral gap, the computer science of undecidability and Hilbert's spectral theory to prove that—like the halting problem—the spectral gap problem was one of the undecidable ones that Gödel and Turing taught us about.

Chatting in that café in Seefeld in 2012, we had an idea for how we might be able to prove a weaker mathematical result related to the spectral gap. We tossed this idea around, not even scribbling on the back of a napkin, and it seemed like it might work. Then the next session of talks started. And there we left it.

A few months later one of us (Toby Cubitt) visited Michael in Munich, and we did what we had not done in Seefeld: jotted some equations down on a scrap of paper and convinced ourselves the idea worked. In the following weeks, we completed the argument and wrote it up properly in a private four-page note. (Nothing in mathematics is truly proved until you write it down—or, better still, type it up and show it to a colleague for scrutiny.) Conceptually this was a major advance. Before now, the idea of proving the undecidability of the spectral gap was more of a joke than a serious prospect. Now we had the first glimmerings that it might actually be possible. But there was still a very long way to go. We could not extend our initial idea to prove the undecidability of the spectral gap problem itself.



BURNING THE MIDNIGHT COFFEE

WE ATTEMPTED to make the next leap by linking the spectral gap problem to quantum computing. In 1985 Nobel Prize–winning physicist Richard Feynman published one of the papers that launched the idea of quantum computers. In that paper, Feynman showed how to relate ground states of quantum systems to computation. Computation is a dynamic process: you supply the computer with input, and it goes through several steps to compute a result and outputs the answer. But ground states of quantum systems are completely static: the ground state is just the configuration a material sits in at zero temperature, doing nothing at all. So how can it make a computation?

The answer comes through one of the defining features of quantum mechanics: superposition, which is the ability of objects to occupy many states simultaneously, as, for instance, Erwin Schrödinger's famous quantum cat can be alive and dead at the same time. Feynman proposed constructing a quantum state that is in a superposition of the various steps in a computation—initial input, every intermediate step of the computation and final output—all at once. Alexei Kitaev of the California Institute of Technology later developed this idea substantially by constructing an imaginary quantum material whose ground state looks exactly like this.

If we used Kitaev's construction to put the entire history of a Turing machine into the material's ground state in superposition, could we transform the halting problem into the spectral gap problem? In other words, could we show that any method

© 2018 Scientific American

for solving the spectral gap problem would also solve the halting problem? Because Turing had already shown that the halting problem was undecidable, this would prove that the spectral gap problem must also be undecidable.

Encoding the halting problem in a quantum state was not a new idea. Seth Lloyd, now at the Massachusetts Institute of Technology, had proposed this almost two decades earlier to show the undecidability of another quantum question. Daniel Gottesman of the Perimeter Institute for Theoretical Physics in Waterloo and Sandy Irani of the University of California, Irvine, had used Kitaev's idea to prove that even single lines of interacting quantum particles can show very complex behavior. In fact, it was Gottesman and Irani's version of Kitaev's construction that we hoped to make use of.

But the spectral gap is a different kind of problem, and we faced some apparently insurmountable mathematical obstacles. The first had to do with supplying the input into the Turing machine. Remember that the undecidability of the halting problem is about whether the Turing machine halts on a given input. How could we design our imaginary quantum material in a way that would let us choose the input to the Turing machine to be encoded in the ground state?

When working on that earlier problem (the one we were still stuck on in the café in Seefeld), we had an idea of how to rectify the issue by putting a "twist" in the interactions between the particles and using the angle of this rotation to create an input to the Turing machine. In January 2013 we met at a conference in Beijing and discussed this plan together. But we quickly realized that what we had to prove came very close to contradicting known results about quantum Turing machines. We decided we needed a complete and rigorous proof that our idea worked before we pursued the project further.

At this point, Toby had been part of David Pérez-García's group at Complutense University of Madrid for more than two years. In that same month he moved to the University of Cambridge, but his new apartment there was not yet ready, so his friend and fellow quantum information theorist Ashley Montanaro offered to put him up. For those two months, he set to work producing a rigorous proof of this idea. His friend would find him at the kitchen table in the morning, a row of empty coffee mugs next to him, about to head to bed, having worked through the night figuring out details and typing them up. At the end of those two months, Toby sent around the completed proof.

IN REMEMBRANCE OF TILINGS PAST

THIS 29-PAGE PROOF showed how to overcome one of the obstacles to connecting the ground state of a quantum material to computation with a Turing machine. But there was an even bigger obstacle to that goal: the resulting quantum material was always gapless. If it is always gapless, the spectral gap problem for this particular material is very easy to solve: the answer is gapless!

Our first idea from Seefeld, which proved a much weaker result than we wanted, nonetheless managed to get around this obstacle. The key was using "tilings." Imagine you are covering a large bathroom floor with tiles. In fact, imagine it is an infinitely big bathroom. The tiles have a very simple pattern on them: each of the four sides of the tile is a different color. You

Tiling an Infinite Bathroom Floor

To connect the spectral gap problem to the halting problem, the authors considered the classic mathematical question of how to tile an infinitely large floor. Imagine you have a box with a certain selection of tiles, and you want to arrange them so that the colors on the sides of each tile match those next to them. In some cases, this is possible by tiling the floor in either a repeating "periodic" pattern or a fractal-like "aperiodic" pattern.

Periodic Tiles

One version of the classic problem concerns tiles that come in three varieties containing five different colors. In this particular case, it is possible to tile the floor with all sides matching up by creating a rectangle that repeats. On each side of the rectangle, the colors match so that many versions of the same rectangle can be placed next to one another in an infinite pattern.

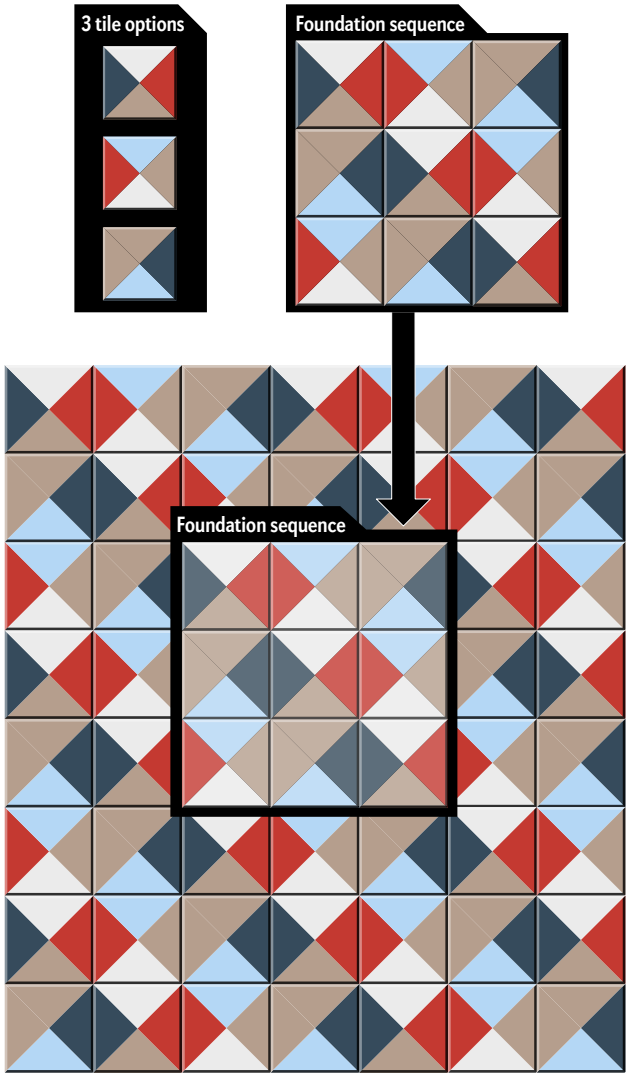
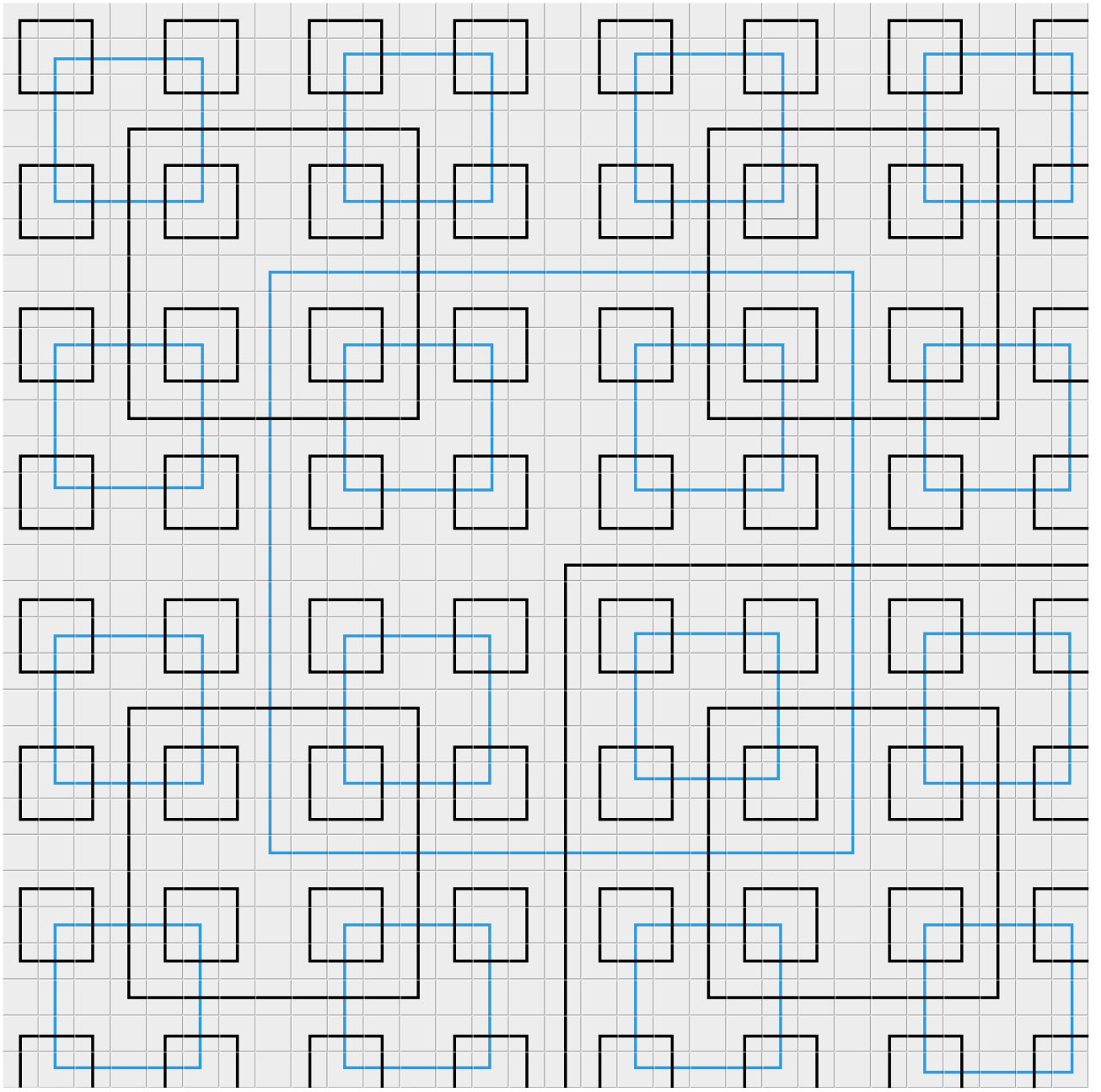


Illustration by Jen Christiansen

© 2018 Scientific American

Aperiodic Tiles

In their proof, the authors used a particular set of tiles designed by mathematician Rafael Robinson in 1971. Robinson's tiles fit together in an ever expanding sequence that does not quite repeat but instead creates a fractal-like pattern. All rotations of the six tiles shown here are allowed. There are also other ways to fit these pieces together in a periodic pattern, but by adding more markings to these tiles (*not shown*), Robinson designed a set of 56 tiles for which no pattern is possible other than the one shown.



© 2018 Scientific American

have various boxes of tiles, each with a different arrangement of colors. Now imagine there is an infinite supply of tiles in each box. You, of course, want to tile the infinite bathroom floor so that the colors on adjacent tiles match. Is this possible?

The answer depends on which boxes of tiles you have available. With some sets of colored tiles, you will be able to tile the infinite bathroom floor. With others, you will not. Before you select which boxes of tiles to buy, you would like to know whether or not they will work. Unfortunately for you, in 1966 mathematician Robert Berger proved that this problem is undecidable.

One easy way to tile the infinite bathroom floor would be to first tile a small rectangle so that colors on opposite sides of it match. You could then cover the entire floor by repeating this rectangular pattern. Because they repeat every few tiles, such patterns are called periodic. The reason the tiling problem is undecidable is that nonperiodic tilings also exist: patterns that cover the infinite floor but never repeat.

Back when we were discussing our first small result, we studied a 1971 simplification of Berger's original proof made by Rafael Robinson of the University of California, Berkeley. Robinson constructed a set of 56 different boxes of tiles that, when used to tile the floor, produce an interlocking pattern of ever larger squares. This fractal pattern looks periodic, but in fact, it never quite repeats itself. We extensively discussed ways of using tiling results to prove the undecidability of quantum properties. But back then, we were not even thinking about the spectral gap. The idea lay dormant.

In April 2013 Toby paid a visit to Charlie Bennett at IBM's Thomas J. Watson Research Center. Among Bennett's many achievements before becoming one of the founding fathers of quantum information theory was his seminal 1970s work on Turing machines. We wanted to quiz him about some technical details of our proof to make sure we were not overlooking something. He said he had not thought about this stuff for 40 years, and it was high time a younger generation took over. (He then went on to very helpfully explain some subtle mathematical details of his 1970s work, which reassured us that our proof was okay.)

Bennett has an immense store of scientific knowledge. Because we had been talking about Turing machines and undecidability, he e-mailed copies of a couple of old papers on undecidability he thought might interest us. One of these was the same 1971 paper by Robinson that we had studied. Now the time was right for the ideas sowed in our earlier discussions to spring to life. Reading Robinson's paper again, we realized it was exactly what we needed to prevent the spectral gap from vanishing.

Our initial idea had been to encode one copy of the Turing machine into the ground state. By carefully designing the interactions between the particles, we could make the ground state energy a bit higher if the Turing machine halted. The spectral gap—the energy jump to the first excited state—would then depend on whether the Turing machine halted or not. There was just one problem with this idea, and it was a big one. As the number of particles increased, the additional contribution to the ground state energy got closer and closer to zero, leading to a material that was always gapless.

By adapting Berger's tiling construction, we could instead encode *many copies* of exactly the same Turing machine into the ground state. In fact, we could attach one copy to each square in Robinson's tiling pattern. Because they are identical copies of the same Turing machine, if one of them halts, they all halt. The energy contributions from all these copies add up. As the number of particles increases, the number of squares in the tiling pattern gets bigger. Thus, the number of copies of the Turing machine increases, and their energy contribution becomes huge, giving us the possibility of a spectral gap.

EXAMS AND DEADLINES

ONE SIGNIFICANT WEAKNESS remained in the result we had proved. We could not say anything about how *big* the energy gap was when the material was gapped. This uncertainty left our result open to the criticism that the gap could be so small that it might as well not exist. We needed to prove that the gap, when it existed, was actually large. The first solution we found arose by considering materials in three dimensions instead of the planar materials we had been thinking about until then.

When you cannot stop thinking about a mathematical problem, you make progress in the most unexpected places. David went over the details of the idea in his head while he was supervising an exam. Walking along the rows of desks, he was totally oblivious to the students working feverishly around him. Once the test was over, he committed this part of the proof to paper.

We now knew that getting a big spectral gap was possible. Could we also get it in two dimensions, or were three necessary? Remember the problem of tiling an infinite bathroom floor. What we needed to show was that for the Robinson tiling, if you got one tile wrong somewhere, but the colors still matched everywhere else, then the pattern formed by the tiles would be disrupted only in a small region centered on that wrong tile. If we could show this "robustness" of the Robinson tiling, it would imply that there was no way of getting a small spectral gap by breaking the tiling only a tiny bit.

By late summer 2013 we felt we had all the ingredients for our proof to work. But there were still some big details to be resolved, such as proving that the tiling robustness could be merged with all the other proof ingredients to give the complete result. The Isaac Newton Institute for Mathematical Science in Cambridge, England, was hosting a special workshop on quantum information for the whole of the autumn semester of 2013. All three of us were invited to attend. It was the perfect opportunity to work together on finishing the project. But David was not able to stay in Cambridge for long. We were determined to complete the proof before he left.

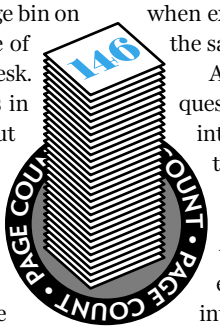
The Isaac Newton Institute has blackboards everywhere—even in the bathrooms! We chose one of the blackboards in a corridor (the closest to the coffee machine) for our discussions. We spent long hours at the blackboard developing the missing ideas, then divided the task of making these ideas mathematically rigorous among us. This process always takes far more time and effort than it seems on the blackboard. As the date of David's departure loomed, we worked without interruption all day and most of the night. Just a few hours before he left for home, we finally had a complete proof.

In physics and mathematics, researchers make most results public for the first time by posting a draft paper to the arXiv.org

© 2018 Scientific American

preprint server before submitting it to a journal for peer review. Although we were now fairly confident the entire argument worked and the hardest part was behind us, our proof was not ready to be posted. There were many mathematical details to be filled in. We also wanted to rewrite and tidy up the paper (we hoped to reduce the page count in the process, although in this we would completely fail). Most important, although at least one of us had checked every part of the proof, no one had gone through it all from beginning to end.

In summer 2014 David was on a sabbatical at the Technical University of Munich with Michael. Toby went out to join them. The plan was to spend this time checking and completing the whole proof, line by line. David and Toby were sharing an office. Each morning David would arrive with a new printout of the draft paper, copious notes and questions scribbled in the margins and on interleaved sheets. The three of us would get coffee and then pick up where we had left off the day before, discussing the next section of the proof at the blackboard. In the afternoon, we divided up the work of rewriting the paper and adding the new material and of going through the next section of the proof. Toby was suffering from a slipped disc and could not sit down, so he worked with his laptop propped on top of an upturned garbage bin on top of the desk. David sat opposite, the growing pile of printouts and notes taking up more and more of his desk. On a couple of occasions, we found significant gaps in the proof. These turned out to be surmountable, but bridging them meant adding substantial material to it. The page count continued to grow.



After six weeks, we had checked, completed and improved every single line of the proof. It would take another six months to finish writing everything up. Finally, in February 2015, we uploaded the paper to arXiv.org.

WHAT IT ALL MEANS

ULTIMATELY what do these 146 pages of complicated mathematics tell us?

First, and most important, they give a rigorous mathematical proof that one of the basic questions of quantum physics cannot be solved in general. Note that the "in general" here is critical. Even though the halting problem is undecidable in general, for *particular* inputs to a Turing machine, it is often still possible to say whether it will halt or not. For example, if the first instruction of the input is "halt," the answer is pretty clear. The same goes if the first instruction tells the Turing machine to loop forever. Thus, although undecidability implies that the spectral gap problem cannot be solved for *all* materials, it is entirely possible to solve it for specific materials. In fact, condensed matter physics is littered with such examples. Nevertheless, our result proves rigorously that even a perfect, complete description of the microscopic interactions among a material's particles is not always enough to deduce its macroscopic properties.

You may be asking yourself if this finding has any implications for "real physics." After all, scientists can always try to measure the spectral gap in experiments. Imagine if we could engineer the quantum material from our mathematical proof and produce a piece of it in the lab. Its interactions are so extraordinarily complicated that this task is far, far beyond anything scientists are ever likely to be able to do. But if we could and then took a piece of it and tried to measure its spectral gap, the material could not simply throw up its hands and say, "I can't tell you—it's undecidable." The experiment would have to measure something.

The answer to this apparent paradox lies in the fact that, strictly speaking, the terms "gapped" and "gapless" only make mathematical sense when the piece of material is infinitely large. Now, the 10^{23} or so atoms contained in even a very small piece of material represent a very large number indeed. For normal materials, this is close enough to infinity to make no difference. But for the very strange material constructed in our proof, large is not equivalent to infinite. Perhaps with 10^{23} atoms, the material appears in experiments to be gapless. Just to be sure, you take a sample of material twice the size and measure again. Still gapless. Then, late one night, your graduate student comes into the lab and adds just one extra atom. The next morning, when you measure it again, the material has become gapped! Our result proves that the size at which this transition may occur is incomputable (in the same Gödel-Turing sense that you are now familiar with). This story is completely hypothetical for now because we cannot engineer a material this complex. But it shows, backed by a rigorous mathematical proof, that scientists must take special care when extrapolating experimental results to infer the behavior of the same material at larger sizes.

And now we come back to the Yang-Mills problem—the question of whether the equations describing quarks and their interactions have a mass gap. Computer simulations indicate that the answer is yes, but our result suggests that determining for sure may be another matter. Could it be that the computer-simulation evidence for the Yang-Mills mass gap would vanish if we made the simulation just a tiny bit larger? Our result cannot say, but it does open the door to the intriguing possibility that the Yang-Mills problem, and other problems important to physicists, may be undecidable.

And what of that original small and not very significant result we were trying to prove all those years ago in a café in the Austrian Alps? Actually, we are still working on it. SA

Toby S. Cubitt is a Royal Society University Research Fellow and reader in quantum information at University College London. After a Ph.D. in physics, postdoctoral positions in mathematics and a faculty position in computer science, he now works on quantum problems that straddle these areas.

David Pérez-García is a professor of mathematics at Complutense University of Madrid and a faculty member at the Institute of Mathematical Sciences in Madrid. He works on mathematical problems in quantum physics.

Michael Wolf is a professor of mathematical physics in the department of mathematics at the Technical University of Munich. His research focuses on the mathematical and conceptual foundations of quantum theory.

MORE TO EXPLORE

Undecidability and Nonperiodicity for Tilings of the Plane. Raphael M. Robinson in *Inventiones Mathematicae*, Vol. 12, No. 3, pages 177–209; September 1971.

Undecidability of the Spectral Gap. Toby S. Cubitt, David Pérez-García and Michael M. Wolf in *Nature*, Vol. 528, pages 207–211; December 10, 2015. Preprint available at https://arxiv.org/abs/1502.04573

FROM OUR ARCHIVES

Ultimate Clocks. W. Wayt Gibbs; September 2002.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

OBITUARY

Life of a Legend

Science mourns the death of physicist and icon of science Stephen Hawking

By Davide Castelvecchi

Stephen Hawking, one of the most influential physicists of the 20th century and perhaps the most celebrated icon of contemporary science, has died at the age of 76. The University of Cambridge confirmed that the physicist died in the early hours of March 14 at his home in Cambridge, England. Since his early 20s, Hawking had lived with amyotrophic lateral sclerosis (ALS), a disease in which motor neurons die, leaving the brain incapable of controlling muscles. Hawking's health had been reportedly deteriorating; just over a year ago he was hospitalized during a trip to Rome.

His death was marked by statements from scientists around the world. Astrophysicist Neil DeGrasse Tyson, director of the Hayden Planetarium in New York City, wrote on Twitter: "His passing has left an intellectual vacuum in his wake. But it's not empty. Think of it as a kind of vacuum energy permeating the fabric of spacetime that defies measure."

One of Hawking's former students at Cambridge, theoretical physicist Raphael Bousso, told *Nature* that his teacher was a brilliant physicist who also excelled at communicating science to the public. "These are two distinct skills. Stephen excelled at both," Bousso remarks.

Now at the University of California at Berkeley, Bousso recalls how he had to learn to shake off his awe and relax around Hawking. "Stephen was a joyful and lighthearted person, not to be burdened by excessively respectful and convoluted interactions," he says.

The British physicist was born in Oxford in 1942. He was diagnosed with ALS when he was 21, while a doctoral student in cosmology at the University of Cambridge. Hawking first realized that something was wrong when he went ice skating with his mother one day, he recalled in a speech on his 75th birthday celebration last year. "I fell over and had great difficulty getting up," he told the audience. "At first I became depressed. I seemed to be getting worse very rapidly."

Although physicians initially gave him just a few years to live, his disease advanced more slowly than expected. He went on to have an active career for decades, both as a theoretical physicist and as a popularizer of science. Still, Hawking progressively lost use of most of his muscles, and for the last three decades of his life was communicating almost exclusively through a voice synthesizer.

Over the years, Hawking became one of the most recognized names in contemporary science. His books, particularly *A Brief History of Time,* became blockbuster successes. He relished making cameo appearances on television shows such as *Star Trek: The Next Generation, The Simpsons* and *The Big Bang Theory.*

But scientifically, his name is most closely associated with the physics of black holes, which he began to study when they were still considered mere mathematical curiosities in Albert Einstein's general theory of relativity. In the early 1970s he began to investigate what quantum physics could reveal about the event horizon, a black hole's surface of no return. Hawking shocked the physics world when he calculated that this surface should slowly emit radiation (soon to become known as Hawking radiation). Black holes were not truly black.

This emission, he reasoned, should ultimately lead the black hole to shrink and disappear. Even more shocking to researchers was the fact that Hawking radiation should erase information from the universe, in apparent contradiction to some of the basic tenets of quantum theory, as Hawking pointed out in 1976.

Perhaps because most of his work was of a speculative nature and difficult to test experimentally, Hawking was never awarded a Nobel Prize. In 2016 some wondered whether he finally might get one, when Jeff Steinhauer, a physicist at the Technion-Israel Institute of Technology in Haifa, announced that he had found convincing evidence of Hawking radiation—not in an actual black hole, but in a laboratory analogue made of extremely cold atoms. Some experts still consider those results inconclusive, however, and many say that their relevance to true black holes is uncertain.

A more direct test of some of Hawking's findings might yet come from the study of astrophysical black holes through gravitational waves, initiated by the U.S.-based Laser Interferometer Gravitational-wave Observatory (LIGO). Hawking and others have linked the surface area of a black hole's event horizon to its entropy, a measure of disorder. When interviewed by *Nature*'s news team in 2016 about LIGO's first detection of gravitational waves from merging black holes, Hawking said that he hoped that future detections would be sensitive enough to confirm a prediction he had made in the 1970: that the surface area of the post-merger black hole should be larger than the combined surface area of the original objects—just as would be expected from entropy, which can never decrease. "I would like them to test my area theorem," he said. SA

Davide Castelvecchi is a senior reporter at *Nature* in London covering physics, astronomy, mathematics and computer science.

MICHELLY RALL *Getty Images*

© 2018 Scientific American

© 2018 Scientific American

ASTRONOMY

Astronomers' newfound ability to see the same cosmic events in light, particles and gravitational waves—*a synthesis called multimessenger astronomy*—gives them a fuller picture of some of the universe's most mysterious phenomena

Messengers from the Sky

By Ann Finkbeiner

Illustration by Maria Corte Maidagan

© 2018 Scientific American

A NEUTRINO HIT ON SEPTEMBER 22, 2017, AT 4:54 P.M. EASTERN TIME. THE NEARLY MASSLESS elementary particle barreled through the sensors of the IceCube Neutrino Observatory, an experiment buried in the Antarctic ice. This neutrino was rare, carrying an energy of more than 100 tera electron volts, about 10 times the energy reachable by particles inside the most powerful accelerators on Earth. Thirty seconds later IceCube's computers sent out an alert with the neutrino's energy, the time and date, and roughly where it came from in the sky.

At the University of Maryland, College Park, IceCube team member Erik Blaufuss got the alert via text and knew that, with that energy, the particle probably came from beyond the solar system. Blaufuss had already seen 10 or so neutrinos in the past year with energies that high, but he thought, "It's a pretty event—let's send it out there." At 8:09 P.M. he issued a public notice over one of astronomy's heads-up networks about the particle, now called IceCube-170922A. IceCube's more than 5,000 sensors, which look for flashes of light made by neutrinos interacting with atoms in the ice, can trace the path of the flash back to the particle's origin in the sky. And Blaufuss hoped the nighttime notice would "catch observers coming online," astronomers who could look at the same area of the sky the neutrino came from. If they were really lucky, they might find the galaxy or other celestial object that sent it.

Neutrinos are just one of the many things in the sky that flare, ping, bang, shudder and shine. For a long time astronomers could see mainly those that shine, that emit light. Then, roughly 30 years ago, they started to detect little hits of neutrinos from beyond our solar system. And since 2015 they have been able to detect the rolling waves of gravity. But combining these different signals to study individual objects—a technique called multimessenger astronomy—is mostly a recent development.

One great advantage of multimessenger astronomy is that unlike light—an electromagnetic wave that can get reflected, absorbed and misdirected, obscuring information about its source—almost nothing stops gravitational waves or neutrinos. The message they carry is pure; it comes in directly and at or near the speed of light. Another plus is that their sources—colliding black holes or collapsing supernovae or merging neutron stars—are transient, unspeakably violent perplexities. They were predicted but not seen, were seen but not understood, or were by any other means invisible. But with more messengers, astronomers can finally understand these complex phenomena. "These sources are complicated," says Francis Halzen, a physicist at the University of Wisconsin–Madison and principal investigator of IceCube. "Unless you have many ways to look at them, you're not going to figure them out."

THE TEXAS SOURCE

FOUR DAYS AFTER BLAUFUSS sent the IceCube notice, scientists at the Swift Observatory's x-ray space telescope reported that since the alert they had counted nine things emitting x-rays in the area of the sky that IceCube-170922A came from.

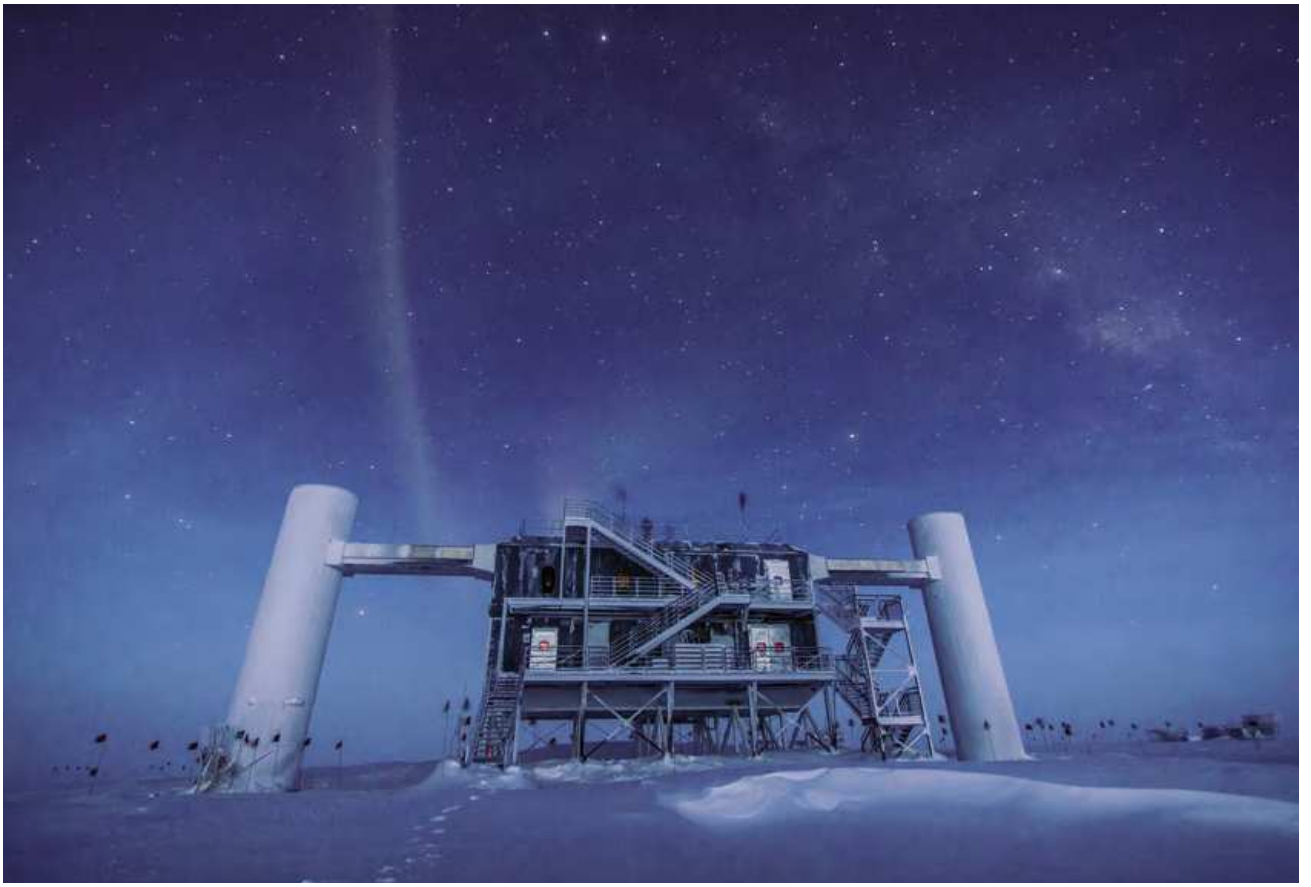
Just two days after that, on September 28, at 6:10 A.M., the Fermi orbiting telescope, sensitive to gamma wavelengths, reported gamma rays at the same position as both IceCube-170922A and Swift's second x-ray source. Sara Buson, a member of the Fermi team at NASA's Goddard Space Flight Center, and her colleagues sent out a public notice saying that the gamma source was already known and named TXS 0506+056, which astronomers later nicknamed "the Texas source." "It was very exciting," Buson says. "The neutrino was exactly on top of the gamma, the first time we had such a nice coincidence." In the previous two weeks, Buson's notice said, Fermi had seen the Texas source flare up by a factor of six.

On the same day at 2:00 P.M., scientists working on a survey called ASAS-SN (pronounced "assassin"), operating at optical

IN BRIEF

Two recent celestial events have ushered in the age of multimessenger astronomy—the technique of observing the same phenomena through light, particles and gravitational waves. **These different messengers** carry unique information, so that combining them gives scientists insight into some of the most mysterious cosmic objects. **Astronomers have traced** gravitational waves and multiple wavelengths of light back to a collision of two dense neutron stars. They have also observed light and neutrino particles coming from what appears to be a huge mass-absorbing black hole.

© 2018 Scientific American



BURIED IN THE ICE at the South Pole, thousands of sensors make up the IceCube Neutrino Observatory. These sensors look for signs of rare interactions between neutrino particles from space and atoms in the ice. A particularly high-energy neutrino observed in 2017 set off multiple observations from ground- and space-based telescopes to identify the particle's source.

wavelengths, announced that the Texas source had in fact been brightening over the past 50 days and was the brightest it had been in several years. The next day, September 29, at 9:00 A.M., another optical telescope found that the Texas source was a blazar, a supermassive black hole at the center of a galaxy that sporadically flares up as matter falls into it, sending out jets aimed straight at us. Then, on October 17, the Very Large Array in New Mexico, operating at radio wavelengths, confirmed the light was coming from a blazar's jet.

Blazars were already well known but had never been observed in multiple wavelengths and simultaneously identified as the source of a neutrino. More interestingly, the Texas source was also the first time a high-energy neutrino coincided in space and time with a similarly high-energy gamma-ray photon. Halzen notes that over the whole sky, the number of high-energy neutrinos and the number of gamma-ray photons are roughly the same, so "the obvious thing is," he says, "it means you could be seeing the same sources." The similarity in numbers, says Imre Bartos, a physicist at the University of Florida, is "a remarkable and suggestive coincidence." But the implication that they are coming from the same cosmic objects, from blazars, Halzen adds, "is a looong extrapolation." Nevertheless, the neutrino discovery could help scientists discriminate between different theories about how blazars manage to accelerate their jets to such energies. "This is a good first step," Bartos says, "but what we need is more multimessenger observations."

A LONG WAIT

THE FIRST MESSENGER that was not light was the neutrino. It came in February 1987 from Supernova 1987A—a dying star whose core collapsed under the weight of its own gravity and then exploded. All in all, scientists detected 25 neutrinos in Japan, the U.S. and Russia. Three hours later optical light came from a shock wave breaking through the star's surface. By November x-rays and gamma rays arrived from decaying radioactive elements and infrared light came from new heavy elements, all created in the explosion. Supernova 1987A helped astronomers understand the way this type of supernova goes off, says Doug Cowen, a physicist at Pennsylvania State University, who is on the IceCube team, and how most of the explosive energy comes out in the form of neutrinos. That was 30 years ago, Halzen says, "and we've been waiting ever since." The coincidence of the IceCube-170922A neutrino with the Texas source—which was eventually observed by at least 19 instruments in gamma rays and x-rays and optical and radio wavelengths—now makes the second neutrino multimessenger event.

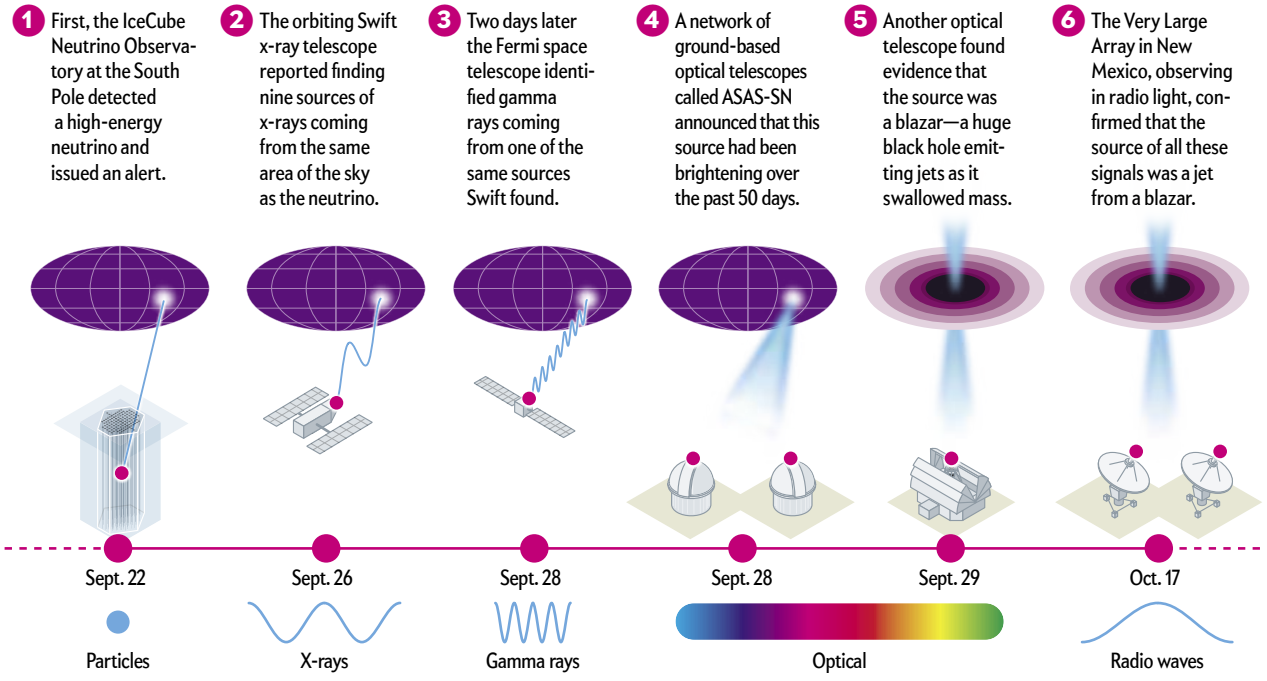
Neutrinos might be excellent messengers, but the most outlandish ones are gravitational waves. These once lived solely in

FELIPE PEDREROS, ICECUBE AND NATIONAL SCIENCE FOUNDATION

© 2018 Scientific American

Cosmic Chain of Events

Over three and a half weeks in 2017, astronomers observed the same celestial event—what they believe to be a flare-up from matter falling into a supermassive black hole—through multiple wavelengths of light, as well as particles called neutrinos. The combined observations offer scientists much more information about these mysterious phenomena than any measurement alone.


1 First, the IceCube Neutrino Observatory at the South Pole detected a high-energy neutrino and issued an alert.
2 The orbiting Swift x-ray telescope reported finding nine sources of x-rays coming from the same area of the sky as the neutrino.
3 Two days later the Fermi space telescope identified gamma rays coming from one of the same sources Swift found.
4 A network of ground-based optical telescopes called ASAS-SN announced that this source had been brightening over the past 50 days.
5 Another optical telescope found evidence that the source was a blazar—a huge black hole emitting jets as it swallowed mass.
6 The Very Large Array in New Mexico, observing in radio light, confirmed that the source of all these signals was a jet from a blazar.
Sept. 22
Sept. 26
Sept. 28
Sept. 28
Sept. 29
Oct. 17
Particles
X-rays
Gamma rays
Optical
Radio waves

the realm of theory, a century-old prediction of Albert Einstein's general relativity. The theory explained the attraction of one mass to another—the apple to Earth—by proposing that mass curved the spacetime around it; the greater the mass, the deeper the curvature. The apple does not so much fall to Earth as it spirals down along the curvature our planet's mass has made in spacetime. The theory went on to predict that if a mass accelerates, the curvature moves outward in waves. The waves are spacetime itself compacting and stretching. So if a gravitational wave moved, for example, through the body of Szabolcs Marka, a physicist at Columbia University, he "would be taller and thinner," he says, "then shorter and wider."

General relativity is widely accepted, and scientists have seen, indirectly, the predicted curvature made by star- and galaxy-sized masses. Gravitational waves themselves, however, had not been seen. In 2014 physicists upgraded an experiment called the Laser Interferometer Gravitational-Wave Observatory (LIGO): two observatories, each with two four-kilometer tubes at right angles to each other. A laser shot from one end of each tube first hits a mirror at the other end, then bounces back, its travel timed. A gravitational wave moving through LIGO would compact and stretch the tubes so that the lasers' travel times would change by one part in 10^{21}, meaning that the four-kilometer tube would be altered by 1/10,000th the diameter of a proton, Marka says, which is like changing the U.S. national debt by one millionth of a cent.

Even with that extraordinary level of precision, LIGO can detect gravitational waves only from extremely dense and massive sources, such as neutron stars: the dropping apple also makes gravitational waves, but comparing an apple's waves to those of a neutron star is not a meaningful exercise. And LIGO's resolution—its ability to locate sources in the sky—is as good as it can be but is still dreadful. With three detectors, one on either side of the U.S. and a third called Virgo in Italy, scientists can trace gravitational waves back to within tens of degrees (the full moon is 0.5 degree across). For an astronomer, says Andy Howell of the University of California, Santa Barbara, that is like waving your hand at the sky and saying, "It's probably somewhere over there."

Between September 14, 2015, and August 14, 2017, LIGO-Virgo detected five different sources of gravitational waves, each produced by the collisions of two black holes that merged into single black holes. These were triumphant observations—the first direct proof not just of gravitational waves but also of the existence of black holes themselves. But they were not multimessenger astronomy. Black holes, being black, are single-messenger events. Current dogma is that they are so dense that light cannot escape them, so their merger is detectable only via gravitational waves. No one expected to see light or neutrinos from these collisions, and although many detectors looked, none did.

COSMIC CRASH

THEN, THREE DAYS AFTER the August 14 black hole merger detection, an event occurred that became a poster child for multi-

Illustration by Nigel Hawtin

© 2018 Scientific American

messenger astronomy. On August 17, 2017, LIGO-Virgo detected gravitational waves. Just 1.74 seconds later the Fermi telescope saw a burst of gamma rays. The event, called GW170817, seemed to be created by the collision and merger not of black holes but of the densest of all stars: neutron stars.

Neutron stars are the collapsed cores of past supernovae, so compact that all their protons and electrons have squished together to make neutrons; they are the final state of stars not quite massive enough to form black holes. The gravitational waves LIGO-Virgo saw would have come from the two stars' inspiral right before they smashed and the gamma-ray burst from their blazing merger—when "all hell breaks loose," says Penn State's Peter Mészáros.

Over the next 24 hours—"like dropping raw meat into a bear pit," says Maryland astronomer M. Coleman Miller—detectors in all frequencies of light on the ground and in the sky scrambled to observe the signal. They pinpointed the merger to a nearby galaxy called NGC 4993 and watched most of its light immediately fade.

The infrared light, however, kept brightening until day three, a sign that as the stars merged, they ejected detritus in which the heaviest of chemical elements were forming. Over the next weeks x-ray and radio light brightened as well, meaning that a near-light-speed jet was pushing through the ejecta. No neutrinos came through, however, so the jets must not have been aimed at us; neutrino detectors "would definitely have seen if it had pointed in our direction," Halzen says. Put together, the messengers were evidence of a phenomenon predicted but never seen, let alone watched in real time: the explosive collision of two neutron stars, called a kilonova. The end stage was either another neutron star, a neutron star on its way to becoming a black hole, or a black hole.

Two months after the kilonova, astronomers announced GW 170817 to the world. That day, October 16, 2017, arXiv.org, a Web site that publishes preprints of science's scholarly papers, received 67 papers. In two months the number of papers roughly doubled: "arXiv is too much," Alessandra Corsi of Texas Tech University says. "I'm having a hard time keeping track."

And just like that, several of astronomy's unsolved problems dropped like swatted flies. The particular kind of gamma-ray burst, a variety that had been seen for decades but whose source had never been directly identified, was now known to come from neutron star mergers. And kilonovae were now understood for the first time to be the birthplaces of a large fraction of the universe's heaviest elements, including platinum, uranium and "about 100 Earth masses in gold," says Samaya Nissanke of Radboud University in the Netherlands. In the weeks afterward, chemists had to rejigger their listings of the sources of the elements on the periodic table. Furthermore, the details of the forms of the gravitational waves cast doubt on the set of alternatives to general relativity proposed to account for the existence of dark matter—possibly excluding the alternative that the universe exists in more than four dimensions.

As always, the finding brought up just as many questions. Astronomers want to know what happens after neutron stars merge. They want to see a neutron star merging with a black hole and to discover how jets arise and what powers them. They still do not know how the cores of stars collapse into supernovae, and they want to watch supermassive black holes in the centers of galaxies merging with other supermassive black holes in the centers of other galaxies.

Thus, aside from the multitude of new and planned detectors of light, scientists envision a whole raft of new multimessenger detectors. LIGO has siblings under construction in Japan and India. The Laser Interferometer Space Antenna (LISA) will be an orbiting gravitational-wave detector scheduled to launch in the 2030s; its arms are lasers zipping among three spacecraft arranged so they form a triangle with sides extending around a million or so kilometers. And new high-energy neutrino detectors are in the works, including a next-generation IceCube and KM3NeT, a cubic kilometer of sensors 3,500 meters down in the Mediterranean Sea.

THE WAGER

MEANWHILE ASTRONOMERS like nothing better than finding things in the sky they were not sure they would ever see. Nissanke got her physics Ph.D. in 2007 from the Paris Institute of Astrophysics, and every day since, she says, she has been thinking about how to see the sources of gravitational waves in light. She would give talks on the subject and would get critical questions. "Astronomers would say, 'You have huge uncertainties, you're measuring tiny displacements, you have huge sky errors.'" When they were not asking questions, they were unimpressed: "Half the audience would look at me like I was on something," Nissanke says. "The other half was asleep." She did this for 10 years.

On August 17, 2017, while speaking at a conference in Amsterdam, she predicted that the first multimessenger events with light and gravitational waves would come in the 2020s. "And the hands went up," she says: "'Samaya, aren't you being overly optimistic?'" After the talk, she had lunch with the LIGO-India consortium, during which she upped her ante: "I don't [usually] bet, but I said I think we'll see the first neutron star merger." Scientists on LIGO-India said not before 2019 and took the wager, a "gentleman/woman's handshake bet," Nissanke says. An hour later LIGO-Virgo saw the neutron star merger. A member of the consortium wrote to her: Before the next conference, let's "tempt nature" and talk about whether we'll see a neutron star–black hole merger.

Nissanke pauses in her story. "I did predict the neutron star's merger, this golden binary, but it took several hours for it to sink in that we were really seeing it," she says. "There's going to be more excitement and many, many, many more papers. It's amazingly fun." SA

Ann Finkbeiner is a science writer based in Baltimore. She specializes in writing about astronomy, cosmology, and the intersection of science and national security. Her most recent book is *A Grand and Bold Thing* (Free Press, 2010) about the Sloan Digital Sky Survey project to map the entire night sky.

MORE TO EXPLORE

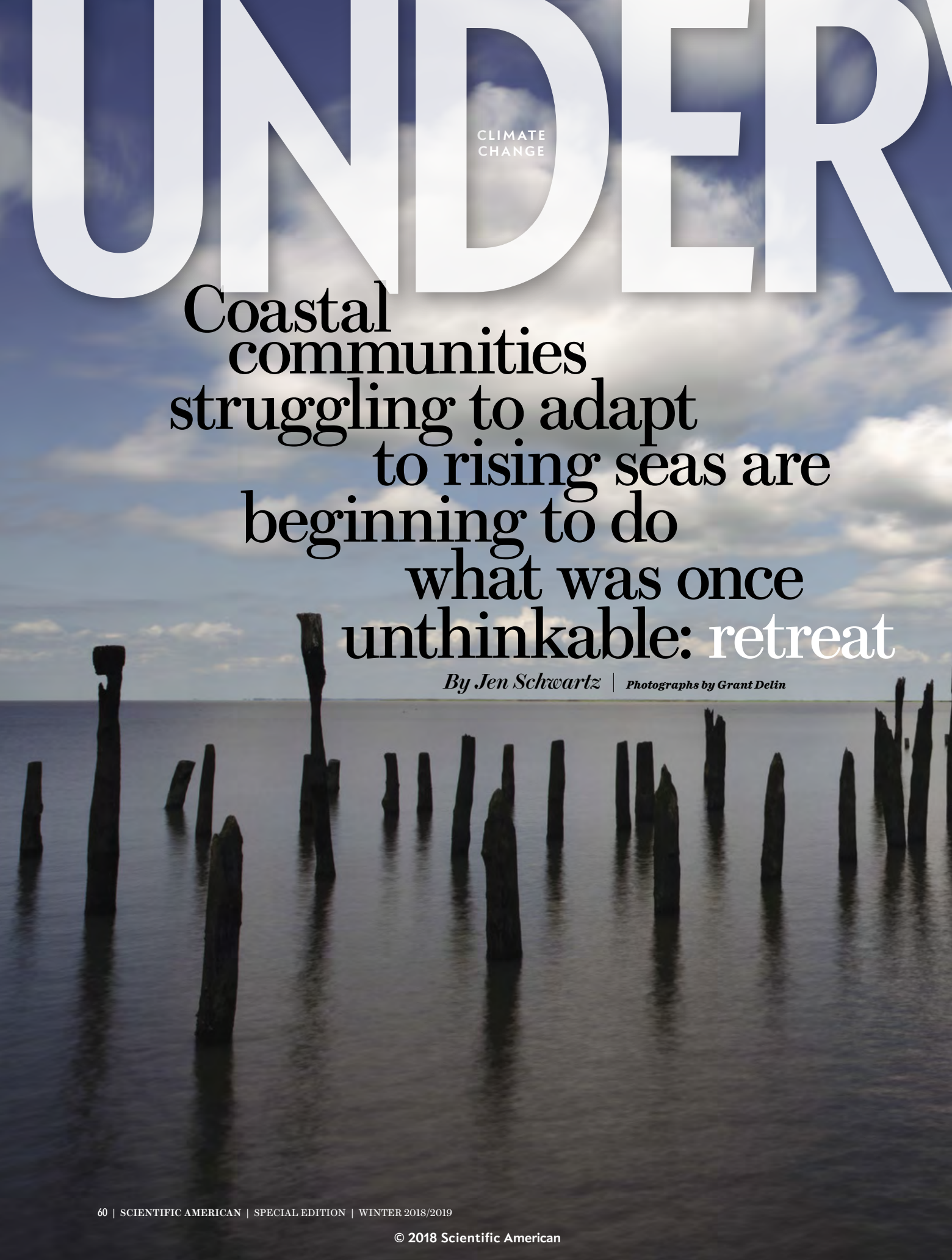
Multi-messenger Observations of a Binary Neutron Star Merger. B. P. Abbott et al. in *Astrophysical Journal Letters*, Vol. 848, No. 2, Article No. L2; October 20, 2017. http://iopscience.iop.org/article/10.3847/2041-8213/aa91c9

FROM OUR ARCHIVES

Neutrinos at the Ends of the Earth. Francis Halzen; October 2015.
Ripples in Spacetime. Clara Moskowitz; December 2016.

scientificamerican.com/magazine/sa

© 2018 Scientific American



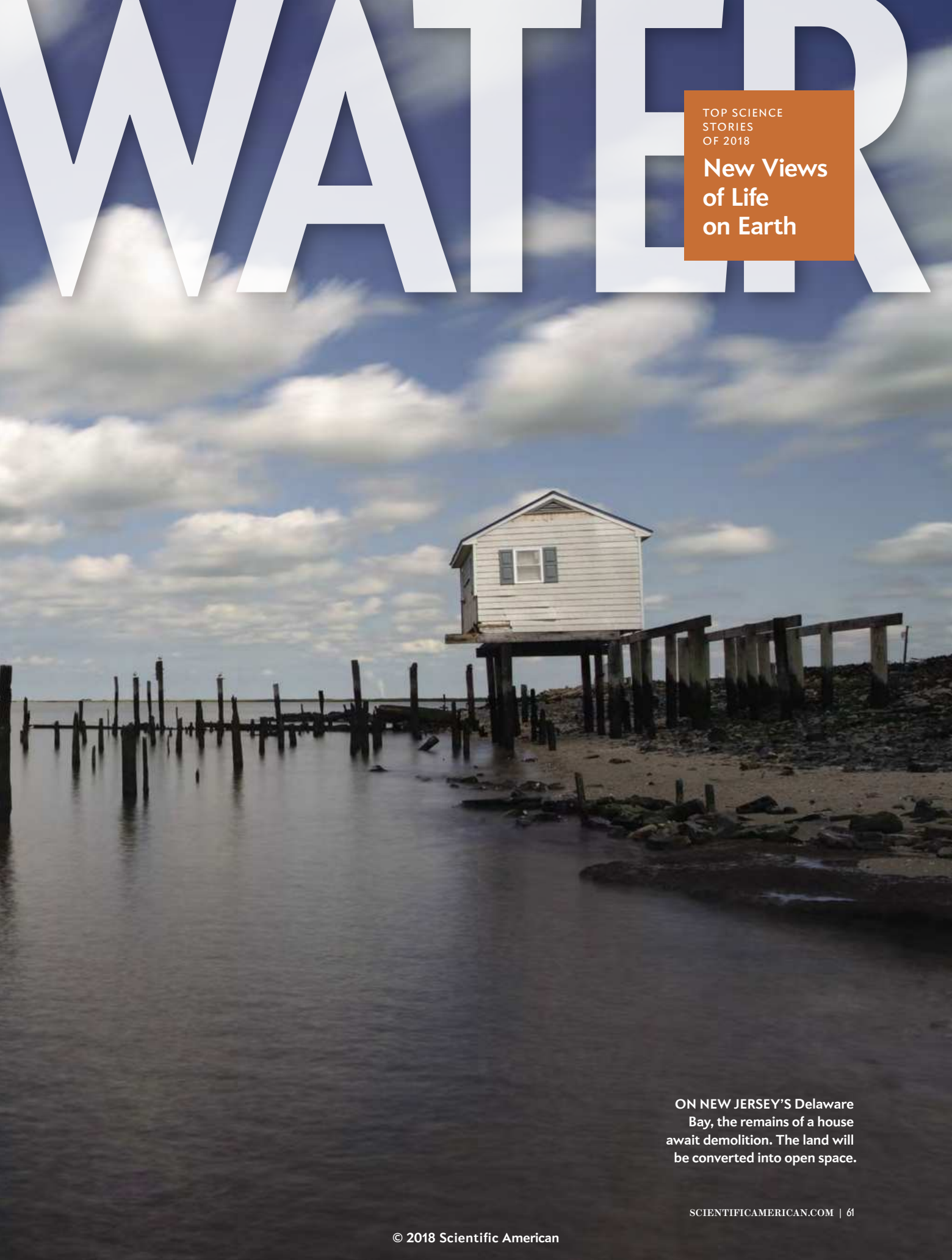
CLIMATE CHANGE

UNDER

Coastal communities struggling to adapt to rising seas are beginning to do what was once unthinkable: retreat

By Jen Schwartz | *Photographs by Grant Delin*

© 2018 Scientific American



TOP SCIENCE STORIES OF 2018

New Views of Life on Earth

ON NEW JERSEY'S Delaware Bay, the remains of a house await demolition. The land will be converted into open space.

© 2018 Scientific American

MONIQUE COLEMAN'S basement was still wet with saltwater when the rallying began. Just days after Superstorm Sandy churned into the mid-Atlantic region, pushing a record-breaking surge into the country's most densely populated corridor, the governor of New Jersey promised to put the sand back on the beaches.

The "build it back stronger" sentiment never resonated with Coleman, who lived not on the state's iconic barrier islands but in a suburban tidal floodplain bisected by 12 lanes of interstate highway. Sandy was being billed as an unusual "Frankenstorm," a one-in-500-year hurricane that also dropped feet of snow. But for Coleman and many residents of the Watson-Crampton neighborhood in Woodbridge Township, the disaster marked the third time their houses had been inundated by floodwaters in just three years. Taxed by the repetitive assault of hydrodynamic pressure, some foundations had collapsed.

As evacuees returned home for another round of sump pumps and mold, Coleman considered her options. Woodbridge sits in the pinched waist of New Jersey, where a network of rivers and creeks drain to the Raritan Bay and then to the Atlantic Ocean. She heard that the Army Corps of Engineers wouldn't be coming to build a berm or tide gate; the area had recently been evaluated, and such costly protections seemed unlikely. Spurred by previous storms, Coleman had already learned a bit about the ecological history of her nearly 350-year-old township. She discovered that parts of her neighborhood, like many chunks of this region, were developed atop low-lying wetlands, which had been elevated with poorly draining "fill" back around the early 20th century. As Coleman researched more deeply, a bigger picture emerged. "I started to realize that, in a sense, we were victims of a system because we were living in a neighborhood that should have never been built," she says.

Although she had flood insurance—her mortgage required it—Coleman knew that her premiums would soon go up, and she worried that her property value would go down. She and her husband liked their house, a prewar colonial. Best of all, it was affordable, a rare find in a town so close to New York City. Coleman had only discovered she would be living in a "special flood hazard area" once she was reading the closing paperwork in 2006. That made her nervous. She recalls her attorney waving it off by saying that at the rate we're going, everyone in New Jersey will live in a floodplain. That might be true in spirit, as a future-looking thought experiment, but it was severely misleading given the circumstances. Desperate to move her family away from a block in Newark with increasing drug activity, Coleman signed away one type of risk for another.

For four uneventful years, the marsh near the bottom of her street was an attractive amenity, a place where her three young sons could play freely. Then the drainages that wrapped around her neighborhood like a wishbone were overwhelmed by a nor'easter in 2010. And by Hurricane Irene in 2011. And again, by Sandy, in 2012.

When federal recovery money started trickling into New Jersey after Sandy, Coleman learned that she could apply for an elevation grant. But raising her house on stilts seemed silly if her car and the road were still on the ground. During Irene, she had witnessed what happens during a storm surge. "The high tide rushes in, and water envelops the entire area in no time at all," she says. "The street becomes a river within a river." Coleman didn't want to be "made whole," in the parlance of disaster-recovery law, if it meant rebuilding in place. Her stress levels spiked every time it rained during high tide. She didn't feel safe, physically or financially.

While commiserating with a neighbor, Coleman heard about a program called Blue Acres. Its premise struck her as radically sensible: The government would "buy out" her repeatedly flooded property at its prestorm value instead of paying to repair it yet again. Demolition crews would then knock down the house and remove other markers of human habitation. She would transfer the deed to the state, and redevelopment would be blocked, forever.

Compared with selling her house, this process seemed overwhelming. But even if she could find a willing buyer, how could she ethically transfer this vulnerability to someone else? "All of us who live in high-risk flood zones were taken advantage of somewhere along the line," Coleman says. "This was a way to end that cycle."

IN BRIEF

As sea levels rise, some coastal communities are confronting a future of both chronic and extreme flooding. Retreat—permanently moving people and property away from the water—is radical, but in some places, it is inevitable.

Even well-managed retreat is messy and painful. It has rarely been tried at all. One New Jersey town is experimenting with moving a cluster of people out of harm's way and turning the newly open land into a flood buffer to protect the rest of the community.

Planning for how and when communities will need to retreat involves many variables. New Jersey could see up to 12 feet of sea-level rise by 2100. Social, political and economic factors add to the complexity.

© 2018 Scientific American

LAND where Monique Coleman's house stood is now planted with saplings. The house to the left will go through the same process.

RETREATING FROM THE COASTS, in concept or practice, is not popular. Why would people abandon their community, the thinking goes, unless no better alternatives remained? To emergency responders, retreat is a form of flood mitigation. To environmental advocates, it's ecological restoration. To resilience planners, it's adaptation to climate change. Everyone agrees, however, that retreat sounds like defeat. It means admitting that humans have lost and that the water has won. "American political institutions, even our national mythology, are ill-suited to the indeterminacy and elasticity of nature," wrote journalist Cornelia Dean nearly two decades ago in her book *Against the Tide*. "It would almost be un-American to concede … that it is we who must adapt to the ocean, not the other way around."

The U.S. has occasionally experimented with retreat on a tiny scale by offering voluntary buyouts to waterlogged families. The outcome is rarely promising. "Buyouts are extremely expensive, extremely disruptive, and many of the attempts have not gone well," says Craig Fugate, former administrator of the Federal Emergency Management Agency (FEMA). They invoke fear among citizens in every political stratum, bringing to mind land grabs, racist resettlement projects, class warfare, and, depending on your ideology, either federal overreach or federal abandonment. Because they require coordination among politicians, homeowners, lawyers, engineers, banks, insurers and all levels of government, they are enormously complicated to execute, even poorly. At their worst, buyouts break up community support systems, entrench inequality and leave a checkerboard of blighted lots in their wake. At their best, they avoid these things and still displace people from their homes.

Yet anyone who has looked at a map that forecasts sea-level rise can see that in low-lying neighborhoods exposed to the tides, some amount of retreat is inevitable. Regardless of how much and how quickly humans cut greenhouse gas emissions, climate change is already producing effects that cannot be reversed. Within a few decades, as saltwater begins to regularly block roads, kill wetlands, disrupt power supplies, bury popular beaches, undermine houses and turn common rainstorms into perilous floods, the most vulnerable pockets of coastal towns will become uninhabitable. As the National Oceanic and Atmospheric Administration has warned, "today's flood is tomorrow's high tide."

Buyouts, however, are not designed for adapting to climate change. Past beneficiaries were almost exclusively riverine communities in the U.S.'s rural interior—people who lived too close to the overflowing Mississippi and Red rivers, for instance, were relocated nearby. The government didn't even begin promoting buyouts as a form of disaster recovery until the 1990s, and since then, they have been conducted as one-off reactions to hurricanes. With multiple federal agencies involved, yet no

© 2018 Scientific American

FAWN McGEE, director of the Blue Acres Buyout Program, has purchased hundreds of houses in New Jersey, including this one. It will soon be demolished.

one taking charge, "it's amazing how much we're still making this up as we go," says Alex Greer, an expert in disaster science at Oklahoma State University. Until recently, retreating from the coasts was practically unheard of.

Superstorm Sandy changed that. The hurricane made sea-level rise, an abstract, future problem for far-flung places, manifest in the form of drowned subway lines and a roller coaster tossed into the waves. It communicated both the experience and evidence of future flooding in a way that probabilities never could. "It's Global Warming, Stupid," said the cover of *Bloomberg Businessweek*. Political leaders in New York State and New Jersey, sensing a tonal shift, realized they couldn't just talk about rebuilding without also talking about resiliency, the rising buzzword of disaster preparedness.

Environmental types were also acknowledging that they could no longer fixate solely on the problem of carbon emissions. Rob Moore was executive director of the Environmental Advocates of New York back in 2012. "We didn't want to talk about adaptation, because we saw it as a distraction from mitigating climate change," he says. "But Sandy made it unavoidable." A few months later Moore took a job at the Natural Resources Defense Council (NRDC) to work on how the country would cope with rising seas. Climate scientists who study the acceleration of sea-level rise felt a similar urgency, and many emerged from their silos to produce better projections. "Now the geophysical people are talking to the atmospheric people and to the economists and the sociologists," says Robert E. Kopp, director of the Institute of Earth, Ocean, and Atmospheric Sciences at Rutgers University and a lead author on major climate reports. This interdisciplinary approach has led to localized forecasting. Instead of only one number—the global mean—we now know that sea-level rise will vary significantly from region to region.

As flooding worsens, a few massive seawalls will likely be built to protect densely populated economic centers, such as lower Manhattan. But there is only so much money, and time, for cement enclosures. Residents in places such as Tangier Island in Virginia and Isle de Jean Charles in Louisiana—and globally from Bangladesh to the Maldives to Senegal—are coping with the same reality as Coleman and her neighbors in Woodbridge Township: a wall isn't coming to save them, and the floods are already here.

NEW JERSEY'S CHIEF OF LAND ACQUISITION clasps her hands on the approach to Bay Point, a peninsula that is disappearing into the Delaware Bay. This is the site of the state's first beachfront buyout, a hard-won success for Fawn McGee. Of the 31 properties she has acquired here, most were recently demolished. Some of the houses were already long gone, leaving behind skeletons of whittled pilings repurposed by nesting ospreys. "It's bittersweet," says McGee from atop a mound of makeshift riprap, residents' last-ditch attempt to stop the erosion. "Even when you can be excited that we bought all these homes, and now we're going to restore the land, everybody is miserable that they had to leave."

When it comes to the unsustainable development of the American coastline, New Jersey owns the honor of being the first and worst. It was here that the tempestuous beach environment,

© 2018 Scientific American

long feared and avoided until industrial times, was rebranded as a summer vacationland. Atlantic City and Cape May were tourist destinations by the mid-1800s, escapes from the undignified swelter and infectious diseases of Philadelphia. The Lenape people had long done the same—a seasonal migration to the shore—but they came to fish, not to conquer the sand with an arsenal of hotels and boardwalks. To hold everything in place, New Jersey was the first state to try to wrest control of coastal sediment flows from nature by erecting seawalls and jetties and bulkheads, and today virtually none of its coastline is untouched by human intervention. It's no surprise that the first speculator to see dollar signs on the sandbar that became Miami Beach was a New Jerseyan.

For the communities McGee works with, retreat has gone from a strategy of last resort to the only option left in the span of about a decade. Climate change drove that shift. But the reason people are in this predicament in the first place is because of the unchecked hubris of coastal engineering, coupled with general human tendencies to love the water and ignore the future. The government gave a huge boost to coastal development in 1968 by introducing the National Flood Insurance Program—an exemplar of moral hazard that allowed homeowners to rebuild over and over in risky areas while keeping their premiums artificially low. Fifty years later that program has accumulated $36.5 billion of debt while effectively trapping people who might prefer to escape to higher ground, NRDC's Moore explains.

Coastal systems, by their tidal nature, are always dynamic and occasionally volatile. The harder we tried to make them stay put, the less stable they became. In the 1960s, when scientists discovered that beaches armored with hard structures actually eroded faster and recovered slower than natural ones, the Army Corps of Engineers began dredging sand from the continental shelf and pumping it back to the shore. Before long, a storm would wash it away, and the dredging would begin anew. Today the Eastern seaboard has a Sisyphean dependence on the "nourishment" cycle. As sea-level rise rapidly accelerates, beach towns are increasingly desperate for fresh infusions of sand, which the corps must travel farther offshore to find. Geologists warn that we are running out of usable sediment faster than the planet can replace it. Wealthy homeowners in Florida are now stealing sand from public beaches to fill in their private ones.

Just as New Jersey pioneered the quixotic development of the coasts, it is leading an acceptance of what it wrought. Twenty-three years ago the state's Department of Environmental Protection (NJDEP) launched the Blue Acres Buyout Program, using state money to purchase a handful of routinely flooded houses here and there. McGee, who runs the program now, was its first leader to leverage federal money from FEMA and other agencies, turning Blue Acres into one of the country's very few standing buyout departments to deal with rising tides. Florida and Louisiana don't even have their own versions. Blue Acres, McGee says, has three major goals: permanently move people and property out of harm's way; open the land for public access; and restore the natural ecology so that the ground becomes a sponge, mitigating flood risk for the rest of the community.

After Superstorm Sandy, McGee prioritized buyout eligibility for entire clusters of houses over individual ones. In that way, the open space created would be big enough to have a real impact on managing floods. She looked for towns based on where enough homeowners had expressed interest in the process *and* the local government was willing to part with a portion of its tax base. After all, when the houses disappear, so, too, goes the tax contribution. Homeowner participation is voluntary, with an ability to pull out at any stage before signing the final contract. That means that families in a "buyout zone" will ultimately have to make a profoundly personal decision about whether or not to uproot in the context of both their neighbors' choices and their town's vision for the future.

The year before Superstorm Sandy, McGee tried pulling off her first large round of buyouts in the aftermath of Hurricane Irene. It didn't go well. "On paper, buyouts make a world of sense, but in practice, that is absolutely not true," Greer, the disaster scientist, says. McGee scrutinized how the process got disjointed and stalled out, making it nearly impossible for anyone to make decisions. It took more than a year for the funding to align, leaving desperate homeowners in a lurch. Some people were upside down on their mortgages, which disqualified them from the program. Or they couldn't afford to keep living in temporary housing while waiting for answers. As disaster amnesia began to set in, many decided just to pursue the readily available methods of rebuilding or elevating. "I realized this went way beyond land acquisition," McGee says. She tasked her tiny Blue Acres team with anticipating the adjacent challenges. "When Sandy hit and the big federal money came in, we were ready," she says.

But advocating for large buyouts of clustered properties meant McGee was pushing for something novel at a moment when New Jersey was in triage mode. The state administration wanted to recover as quickly as possible. "I think they looked at me as a troublemaker for challenging us to think beyond black-and-white," she says. McGee was arguing to disrupt the cycle of hasty rebuilding in areas where she knew the water would rise again. After many persistent meetings (her approach was "confidence, not confrontation," she says) and assurances that she could navigate through the bureaucracy, McGee convinced the state to give her plan the go-ahead.

Immediately McGee asked to borrow 33 staffers from other areas of the NJDEP. "In addition to my GIS [geographical information system] people, I brought in four paralegals, six project managers to do grant writing and deal with the feds, and eight case managers to walk families through the process," she says. "Then title people. Appraisers. Surveyors. Accountants. We had bookkeepers just to go through shoeboxes of Home Depot receipts from homeowners and cross off things like Snickers bar, water, couch." With each snag that threatened the buyouts, McGee found herself filling some unexpected role. She played a therapist to storm victims experiencing trauma, illness and divorce. When McGee saw how many homeowners were upside down on their mortgages, she facilitated short sales that amounted to more than $5.4 million of debt forgiveness—which resulted in 67 additional bought-out homes. When the issues went beyond her scope, she outsourced. "Hoarders! They literally can't get out of their houses," she says. "We had to bring in specialists to help them let go of their stuff."

Most important, McGee sensed that buyouts were a deeply communal decision. She figured neighbors would be looking at their options together over a bottle of wine, so she assigned case managers according to social groups. Case studies of buyouts show that individual considerations play only a limited role in whether a family leaves or stays. "The implication here is that residents'

© 2018 Scientific American

Flooding Up Close

When people think about sea-level rise (SLR), the first thing that comes to mind is the beach—narrow, sandy environments that are visibly exposed to surging water and powerful waves. Although New Jersey is known for its Atlantic coastline, much of it composed of barrier islands that are engineered to stay in place, it has a tidally influenced shoreline spread across 20 of its 21 counties. As SLR accelerates through this century, it will make tidal flooding more common and storm flooding more extreme, causing economic, environmental and social upheaval in all types of tidal communities, from bay-shore wetlands to barrier islands to suburbia. These are not problems for the next generation. According to a June 2018 report by the Union of Concerned Scientists, 250,000 homes in New Jersey, together valued at $108 billion, are at risk of chronic flooding within the next 30 years.

FROM A DISTANCE

These maps show what areas will likely be inundated by high tide at 2050 and 2100. Whether we severely cut emissions or keep polluting as usual matters little for 2050 projections; existing effects aren't reversible. (The choices we make today *will* have a big difference on how much the seas swell after 2050.) Another major factor for New Jersey is subsidence: the land itself will sink by about 0.4 feet by 2050. Relative to the baseline sea level from 2000—the average high tide from 1991 to 2009—the New Jersey coast is likely to see one to 1.8 feet of rise by 2050, according to research by Rutgers University climate scientist Robert E. Kopp and his colleagues. There is a 50 percent chance that high tide will reach or exceed the median projection of 1.4 feet above the 2000 baseline within three decades.

New Jersey Sea-Level Projections

2050

- 1.4 feet of sea-level rise (high emissions, "business as usual")

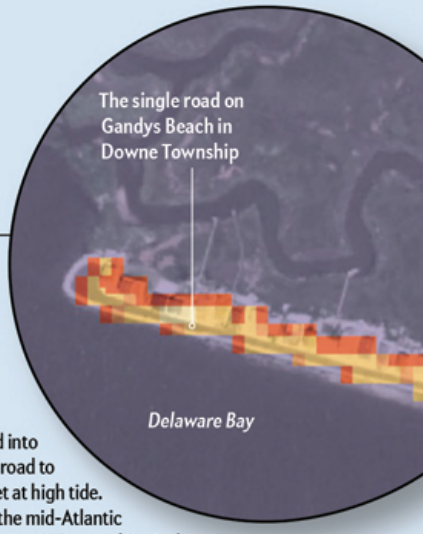
2100

- Lower emissions; more stable Antarctic
- Lower emissions; less stable Antarctic
- High emissions; more stable Antarctic
- High emissions; less stable Antarctic

New Jersey

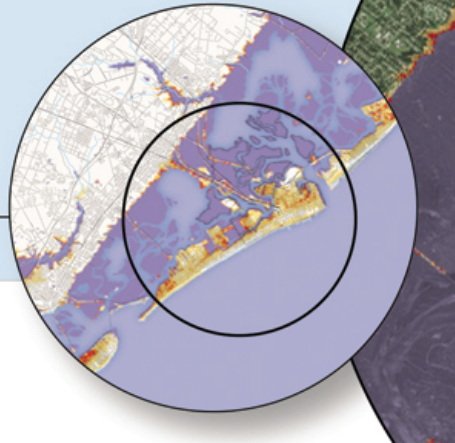
GETTING LOCAL

Zoom in, and the range of near-future threats becomes clear. Houses will routinely flood. High tide will cut off intercoastal access roads. Popular tourist beaches and protective wetlands could disappear. And these higher water levels will be intensified by every full moon, rainstorm, nor'easter and tropical storm. "Increasingly, the research is based on compound threats of flooding instead of the single factor of sea-level rise," says Bryan Jones, a geographer who models climate migration.



BAY WETLANDS

Gandys Beach hosts a single row of homes propped up on pilings, some of which are now always over the water. In May 2018 a portion of the island's road had collapsed into the waves, and the access road to the mainland is already wet at high tide. Many coastal wetlands in the mid-Atlantic region will not likely survive an SLR rate of 0.4 inch a year, which could occur here between 2030 and 2040. Without these natural buffers, communities lose one of the most effective protections against storm surges and tidal flooding.



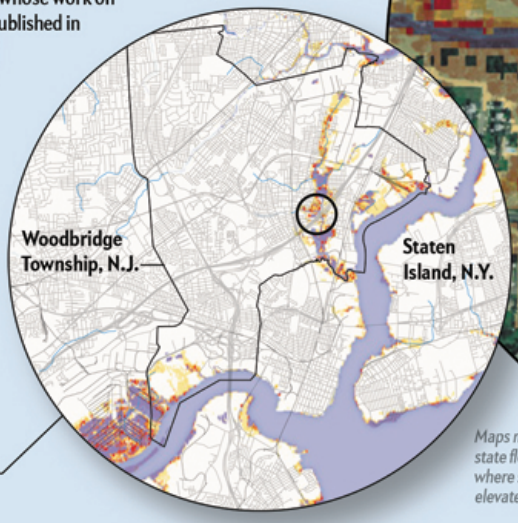
Map by Mapping Specialists, Graphic by Jen Christiansen

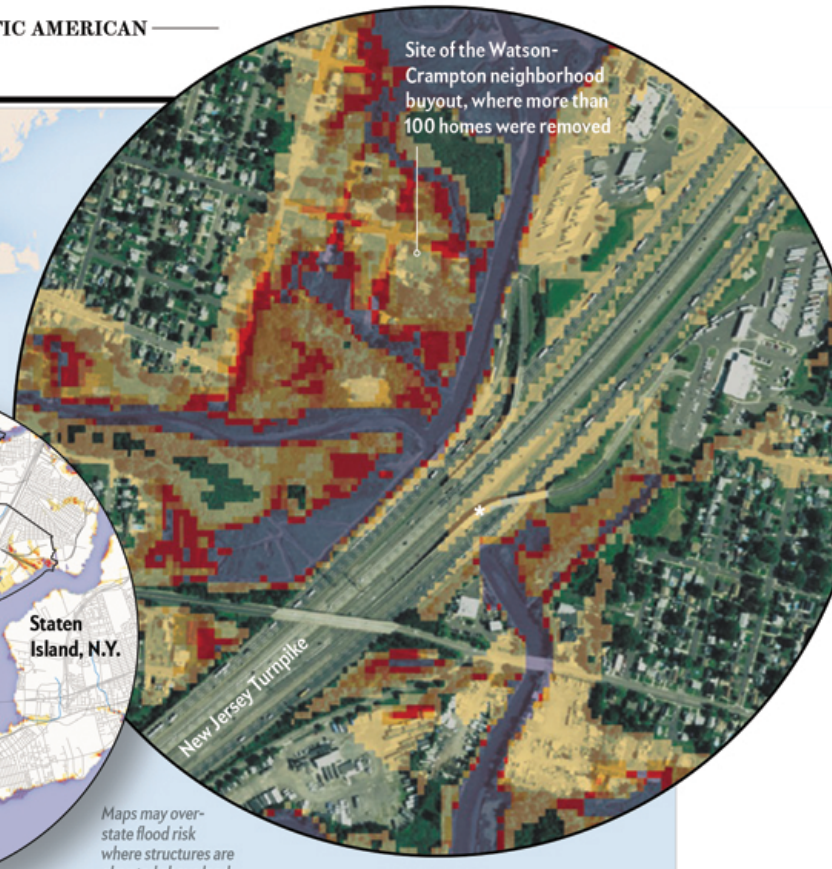
SOURCES: CLIMATE CENTRAL (*base flood maps*); "EVOLVING UNDERSTANDING OF ANTARCTIC ICE-SHEET PHYSICS AND AMBIGUITY IN PROBABILISTIC SEA-LEVEL PROJECTIONS," BY ROBERT E. KOPP ET AL., IN *EARTH'S FUTURE*, VOL. 5, NO. 12; DECEMBER 2017 (*sea-level projections*); ROBERT E. KOPP (*flood height*); U.S. DEPARTMENT OF AGRICULTURE (*base satellite imagery*)

© 2018 Scientific American

URBAN FLOODPLAIN

Dense residential areas such as Woodbridge Township are especially vulnerable as their tidal waterways swell. "When increased river discharge, precipitation and storm surges all happen at the same time, coastal regions often experience flooding that is much, much worse that just storm surge or river flooding on their own," says oceanographer Thomas Wahl, whose work on so-called compound flooding was published in *Nature Climate Change* in 2015.

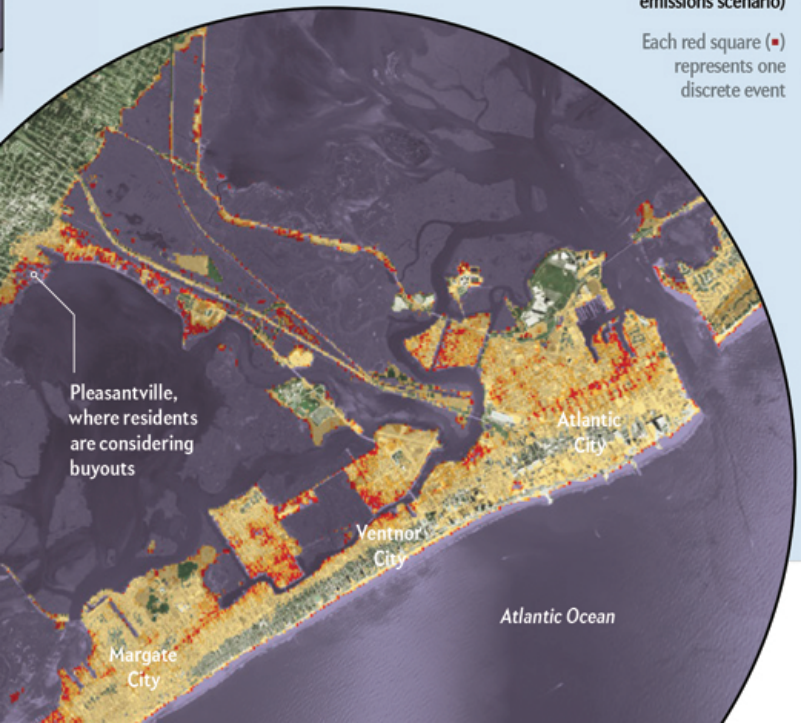




Maps may overstate flood risk where structures are elevated above land

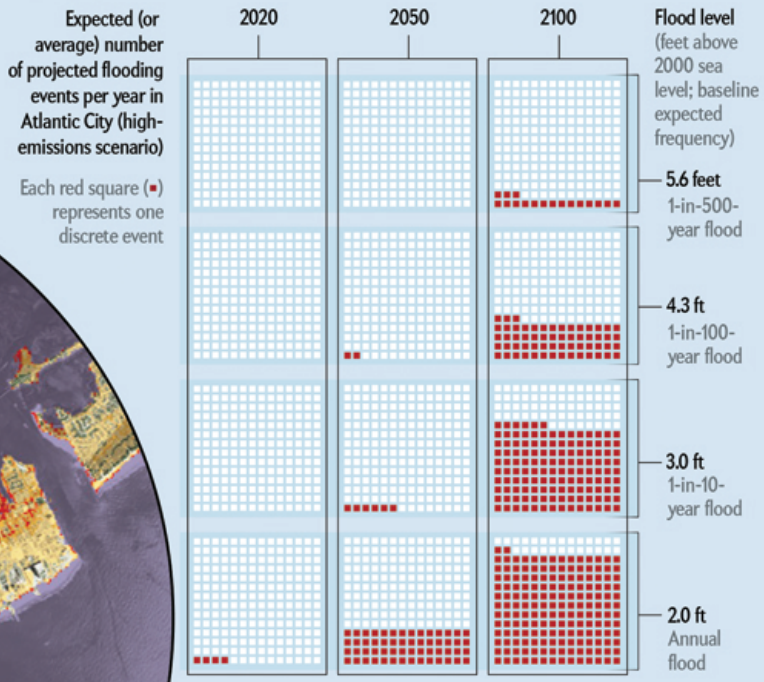
BARRIER ISLANDS

Long before popular beach towns such as Atlantic City, Ventnor City and Margate City (*shown below*) are permanently drowned, the effects of climate change will dismantle them in other ways. "Most people perceive the value of the beach as the sand itself," says Craig Fugate, former head of FEMA. When there is not enough space between the built environment and the water's edge to host sunbathers, will housing prices plummet? Without a sandy beach, will the town itself still have an economic raison d'être? "The question for the Jersey Shore is, Do you start thinking about giving up that first row of homes, businesses and streets in order to get the beach back?" Fugate asks. "Or will we climb up a seawall and walk directly into the water?"



A FUTURE OF REGULAR FLOODING

The maps here show projections of how the baseline sea level will change in the future, inundating currently dry land during high tide. Just because a certain block or property isn't getting wet every day doesn't mean it will be desirable, or even possible, to live and work there. Tidal flooding, sometimes called "nuisance" or "sunny day" flooding, will occur regularly because the baseline water height is getting higher. And when storms arrive, their surges and waves will be more destructive and far-reaching. This chart shows how the frequency of flood events at a given height in Atlantic City are projected to increase as baseline sea level grows.


Expected (or average) number of projected flooding events per year in Atlantic City (high-emissions scenario)
Each red square (▪) represents one discrete event
2020
2050
2100
Flood level (feet above 2000 sea level; baseline expected frequency)
5.6 feet 1-in-500-year flood
4.3 ft 1-in-100-year flood
3.0 ft 1-in-10-year flood
2.0 ft Annual flood

© 2018 Scientific American

perceptions of their broader community were more influential in the buyout decision than were their experiences of Sandy, or their personal or family characteristics," wrote Sherri Brokopp Binder, an expert on postdisaster relocation, and her colleagues in a 2015 study in the *American Journal of Community Psychology*. McGee realized that grassroots support from a few residents could make all the difference between a transformative buyout and a dud.

IT DIDN'T TAKE LONG after Sandy for Monique Coleman to decide that she wanted a buyout. Now she would have to convince her neighbors and the town to want the same. McGee remembers meeting Coleman in 2013, when she hosted a standing-room-only meeting in Woodbridge to tell the residents about Blue Acres. "People would listen to me but then look over at Monique for her to signal it was okay," McGee says. "We always have police at these meetings because people get heated, but Monique had this calming effect on her neighbors. It was astonishing."

Coleman, it turned out, had exposure to grassroots organizing dating to her childhood. Her grandmother took her to Black Liberation Movement meetings, and her father brought her door-to-door during his city council run. These days Coleman works as a teacher for blind children, so in a sense her profession is helping others adapt to unfamiliar challenges. When Coleman and a handful of buyout supporters began canvassing their own streets in late 2012, she expected resistance. "You can't present this idea once, and when you hear that 'no,' just say, 'Okay, I understand,'" she says. She set up a blog and a Facebook group and organized monthly meetings. It was going to take a lot of listening at a time when people were traumatized by Sandy and overwhelmed by the day-to-day choices of recovery.

Blue Acres has been touted as a model for how retreat might evolve. Roy Wright, former resilience lead at FEMA, called New Jersey's approach to buyouts "masterful." But in all McGee's meetings and maps and her talks with mayors and residents, she was not able to evoke the language or data of climate change. Until January 2018, under the administration of Governor Chris Christie, McGee "was not to use those terms," as she puts it.

So Coleman took it on herself to "learn the scientific side of things." When she started doing Web searches for the impacts of climate change in 2012, she couldn't find much that spoke to her area specifically—it simply didn't exist yet. But what she did find convinced her that the back-to-back floods she had experienced in Woodbridge weren't a spate of bad luck but signals of a new reality. Coleman wrote and distributed a leaflet of 12 reasons why life in the flood zone wasn't sustainable in a climate-changed future.

Eventually some of her neighbors stopped slamming their doors when she showed up to chat about buyouts and started asking her questions instead. Coleman was patient but persistent. "It is very hard for people to accept this is really happening," she says. She explained that no matter how she sliced it, she found no positive paths forward. Leaving for good was simply the least onerous. "Who knows when a flood will come again," she would tell her neighbors. "But it will be coming."

THE MODELS SCIENTISTS HAVE MADE to predict the influence of sea-level rise on those future floods have become highly sophisticated, combining global factors such as the thermal expansion of the seas with local variables such as land subsidence and variations in the gravitational pull of land on the ocean around it. But major uncertainties remain. For one thing, we don't know how quickly and severely societies will cut greenhouse gas emissions. For the next few decades certain effects will occur regardless of how much we mitigate climate change. Rutgers's Kopp, a leading climate scientist, says that New Jersey will likely experience between one and 1.8 feet by 2050. Even at the low end, numbers like that will reshape life along the coast. After 2050 the rise continues to accelerate, but the picture gets murky: NOAA estimates that New Jersey could see between three and 12 feet of sea-level rise by 2100. That range is overwhelming if you're a mayor who is trying to come up with an adaptation strategy. "Climate change is a study of probabilities, but the public wants yes or no answers," says Graham Worthy, director of the National Center for Integrated Coastal Research.

Besides the human element, however, the biggest wild card when it comes to the fate of coastal communities in New Jersey and beyond is the stability of the West Antarctic ice sheet. The rate at which the whole of Antarctica is shedding ice has tripled over the past decade, and West Antarctica, being especially sensitive to the forces of climate change, is one of the most rapidly changing places on the planet. The West Antarctic ice sheet is so voluminous that it will add more than 10 feet of sea-level rise alone if it catastrophically collapses. This scenario might only be avoided with extreme emissions cutbacks in the next decade, according to a June 2018 report in *Nature*.

NASA's satellites have been collecting data on this ice sheet and others for more than four decades. Satellites are unrivaled at capturing continuous observations over a wide area, but they can't pick up details that would make it possible to predict the ice sheet's fate with a higher degree of certainty. Some of those more granular clues include ice thickness, the grounding line where a glacier's base meets the sea and the slope of the ice sheet, which is a driving force that sends ice from the interior of the continent to the ocean, says Kenneth Jezek, a glaciologist and retired polar researcher. Because of its sheer size and remoteness, studying West Antarctica up close is a dangerous, logistical nightmare.

One survey project, a NASA airborne mission called Operation IceBridge, has been able to capture some of those details by flying above the region in a retrofitted jetliner. From an altitude of just 1,500 feet, scientists onboard this winged laboratory can see that the top of the ice sheet is textured with signs of movement, such as geometric crevasses, the milky cerulean of ancient ice exposed to sunlight and cracked-up plains that resemble lake beds in drought. These features, they know, are carved by katabatic winds from above and by invisible rivers from below. But it is the ridges of black bedrock that hint at dramatic topography underneath the ice—a hidden world that IceBridge has been mapping extensively over the past decade.

To understand what is happening under the frozen surface, pilots maintain a precise track over the sheet while radar streams data to an onboard computer screen, revealing evidence of entire mountain ranges and valleys that make up the shape of the continent. A gravimeter picks up the depth and size of glacial cavities filled with seawater, a marker of how floating ice shelves might be melting as they interact with the ocean. Along the glaciers' terminal edges, icebergs float against the inkiness of the Amundsen Sea, a scene photographed every second by two cameras affixed to the belly of the plane. IceBridge has flown some of these tracks year after year, capturing change in unprecedented

© 2018 Scientific American

AS HOUSES come down, Thomas C. Flynn, Brooke Maslo and Jeremiah Bergstrom are restoring floodplains in Woodbridge Township, N.J.

detail. At a time when the necessity of earth science is being attacked by political leaders, "I can't emphasize enough that we do not collect this data because we find it scientifically interesting," says John Sonntag, IceBridge's mission scientist. "We collect it to try to warn and protect our communities from sea-level changes that are coming their way."

As raw data from IceBridge, satellites and similar projects have percolated through scientific papers and reports such as the 2017 National Climate Assessment, new tools have emerged. NOAA's Digital Coast and Climate Central's Surging Seas, for instance, allows town planners to begin envisioning how sea-level rise will affect flooding in their jurisdictions.

IceBridge data have turned out to be essential for filling in fundamental gaps in polar ice knowledge, "but we still have a way to go in Antarctica," says Eric Rignot of the University of California, Irvine, and NASA's Jet Propulsion Laboratory. Rignot was the lead author on a landmark 2014 paper in *Geophysical Research Letters* that used radar measurements of a large sector of West Antarctica and concluded that it "is undergoing a marine ice sheet instability that will significantly contribute to sea level rise in decades to centuries to come." That same week a *Science* paper suggested that, based on modeling, the collapse of the West Antarctic ice sheet had already begun, making extreme sea-level rise inevitable, possibly within two centuries. But Rignot thinks that time line might be too conservative. Real observations of ocean temperature—and how glaciers are responding to those warmer waters—are still "totally lacking" in parts of Antarctica, meaning that "it is a matter of fact that our projections tend to underestimate sea-level changes," he says.

Neither Rignot nor Jezek thinks that the next NASA ice-monitoring satellite, which launched on September 15, will be enough to narrow the uncertainties coming out of Antarctica. Rignot suggests that it will take more airborne research such as IceBridge, along with shipborne surveys by unmanned subs and multibeam sonars and new forms of intelligent technology—"an entire army of robotic devices"—dispatched to the remotest fringes of the continent.

Next up, science teams led by the U.S. and U.K. will travel by air and icebreaker to West Antarctica's marquee feature—the massive and infamously unstable Thwaites glacier—to do just that. More than 100 scientists from around the world will study the interaction between warming ocean water and the ice shelf to examine how Thwaites is thinning from below. Thwaites is like a tub stopper holding much of West Antarctica in place; if it's doomed, so is the ice shelf. The more these researchers and others learn about the shifting dynamics among ice, ocean water and atmosphere, the more factors they can plug into regionally specific sea-level predictions. The data they gather will inform whether coastal populations have centuries, or mere decades, to prepare for the onset of the deluge.

BY THE SUMMER OF 2014, when Coleman signed over the deed to her old house and moved into a new one, Woodbridge Township was on its way to becoming the site of Blue Acres' biggest ever buyout project. Today a total of 142 homeowners have accepted offers. And about 115 homes from the Watson-Crampton

© 2018 Scientific American

TIDALLY INFLUENCED wetlands abut the Watson-Crampton neighborhood in New Jersey. By 2050 it will likely be underwater.

neighborhood alone have been removed, most of them clustered within a grid of streets covering about 30 acres. Millions of people globally will have to move inland to escape the coming floods, so these numbers can sound too tiny to be meaningful. But what happened in Woodbridge upends many of the assumptions traditionally tied to buyouts: that no residents want to leave, that politicians will never get onboard, that ecological health in suburbia will never win out over real estate growth and that no one is planning for a climate-changed future by making painful choices in the present. "What we're doing here is paving the way for conversations about how to manage retreat," says Thomas C. Flynn, the town's floodplain manager.

Officially the buyout process ends once the property has been demolished. Woodbridge, however, found itself with an abundance of lots and ambitions beyond mowing the grass. The town reached out to Rutgers ecologist Brooke Maslo, who works with the school's Cooperative Extension to assist New Jersey communities with science-based projects. The term "resilience" gets tossed around a lot, "but what does it actually translate to?" Maslo asks. She came to see the Watson-Crampton buyout project as a unique opportunity: she could create a floodplain restoration that buffered the remaining neighborhood from sea-level rise. She brought in Jeremiah Bergstrom, a landscape architect with experience managing stormwater in urban environments. "As far as I can tell, this is the first coastal land restoration in the context of residential retreat," Bergstrom says.

Using nature as infrastructure is a well-established concept—think mangroves and oyster beds as storm-surge absorption—but it is not commonly applied in places as densely developed as the greater New York City area. Liz Koslov, an assistant professor at the University of California, Los Angeles, who did ethnographic research on Staten Island's post-Sandy buyouts, says she has seen next to no discussion about what happens to the land itself after the houses come down. "Residents said they just want the land to go 'back to nature,' but when you get down to it, 'nature' can mean a lot of different things," she says. Karen O'Neill, a Rutgers sociologist who is cataloguing global instances of coastal retreat, says that "you hardly ever see a comprehensive ecological restoration. It just doesn't exist."

The Watson-Crampton neighborhood can't simply return to nature, because it was built on fill. "We have to re-create a new ecology, a new nature," Bergstrom says. Over the past year the restoration team has ripped out roads, assessed soil quality, and planted more than 950 saplings to increase flood storage capacity and encourage the growth of a biodiverse salt-marsh habitat. Without intervention, the land would become a monoculture of invasive reeds that can break down and form dense mats, which might ultimately make flooding issues worse, Flynn explains.

© 2018 Scientific American

Maslo and her team are softening the hard curvature of the swale so that tidal surges entering from the river below the New Jersey Turnpike don't rush in at high velocity. They'll carve a channel that will allow for spillover, with the hope that the depressions will create permanent standing water for wildlife. Maslo wants to prove that a town can recoup its tax losses with new lures, such as parkland trails and a kayak launch. "This is not a wasteland," she says.

Maslo's vision helped to convince the Woodbridge mayor and city councillors to change the township's buyout regions—120 acres in all—from residential zoning and existing marshland to something they named the Open Space Conservation/Resiliency Zone. No development would be allowed. Nineteen households in the Watson-Crampton buyout area dug in their heels. The town warned these "holdouts" that if they ever wanted to sell, their houses would first need to meet new floodplain standards—which would likely mean elevating them higher. Land-use changes like this are controversial because they make what is supposed to be a voluntary process into one that is significantly less so. But without them, developers might be attracted to come in and build bigger and higher properties. Then new people—those who can afford higher flood insurance premiums and the building costs of living in a floodplain—are likely to move in, replacing those who can no longer afford to stay.

FOUR YEARS AFTER RELINQUISHING her home to the forces of nature, Coleman says she has no regrets about taking the buyout. For moving within the same county, she received an incentive grant of $10,000, which helped her afford another single-family home on higher ground. The process was financially and emotionally stressful, but the way Coleman describes her participation reframes a reaction to misfortune as a deliberate act. In an era of climate refugees who will be forcibly cast out of their homes by either too much water or not enough, Coleman sees herself as more of a retreat pioneer—someone who seized whatever agency she could as she faced an uncertain future. "There's nothing worse than sticking your head in the sand and resisting all this change going on around you," she says. "Because then you end up feeling pushed to make a decision that you are not prepared to make."

McGee, meanwhile, is playing the long game. In the spring of 2018, five and a half years after Sandy, Blue Acres was still submitting new buyout applications to the federal government. Woodbridge is on its third round of buyouts, which involves seven of the 19 holdouts that remain in the resiliency zone. "I don't close out a grant until we've done enough demolition so that the holdouts can digest how the character of their community is changing," McGee says. The tactic is working. She has spent $172 million of her total funding pot, which has grown as other recovery programs failed and FEMA directed the unused money her way. Blue Acres has facilitated nearly 1,000 buyout offers since Sandy, of which 713 have been accepted by homeowners. "You think it'd be 10,000 families for all the work we've done, but it's not, because it's so damn hard," McGee says.

Woodbridge's resiliency zone is not quite ready to be held up as a demonstration project. The work isn't done, and the marsh isn't yet beautified. Over the years, as the houses came down in stages, the scene sometimes looked like an eerie abandonment of the built environment, not a harbinger of progressive adaptation. But now that nearly all the structures are gone and the outlines of formerly paved roads are blurring into grass, "it looks less like a ghost town and more like it's just land," says Coleman, who visits every few months to see the transition in progress. "Now it's the houses that look like they don't belong."

No one disagrees that undeveloping certain areas of the coastline will be messy and expensive. But as the science of coastal resilience becomes more collaborative, the how-to of retreat may become less daunting. Retreat, after all, will not mean drawing a line some distance inland from Maine to Florida and removing everything to the east, explains Bryan Jones, a geographer who models climate-induced human migration. Modeling combined with artificial intelligence is now producing tools that allow planners to play out what-if scenarios in their towns. If, say, you buy out 40 houses from one location, restore groundwater storage and run the 100-year flood of the future, would that significantly reduce damage to adjacent houses? What are the social and economic trade-offs of undeveloping one neighborhood to protect another? Can land be designated as a safe relocation spot? These are the kinds of questions that Fugate, who led FEMA during Sandy, and others are working on now. "Just as the quantification of catastrophe risk drove a huge expansion in catastrophe insurance, it is about to drive a great industry of disaster risk reduction," wrote Robert Muir-Wood in his 2016 book *The Cure for Catastrophe.*

Retreat is so new that few planners are thinking about the next step: relocation. "Globally, there's substantial evidence that people end up right back in harm's way," Jones says. In a survey of Staten Island families who took buyouts after Superstorm Sandy, Binder, the sociologist, found that 20 percent moved to a home that is equally or more vulnerable to flood risk. As more people begin to flee slightly inland, they will encounter a wave of people still moving toward the coast. Just like backwash hitting the surf, the result could be turbulent.

As U.C.L.A.'s Koslov wrote in 2016 in *Public Culture,* "the complexity and ambivalence of retreat serves as a reminder that there are no easy solutions and that it is not possible to rebuild forever or to wall ourselves off from the problems we face." Retreat signals not just the physical movement of recalibrating to the tides but an existential reckoning with our ways of living along the water. The word itself is borrowed from the language of geologic processes, which humans have undeniably hastened. As glaciers and beaches retreat, so, too, will we. SA

Jen Schwartz is a senior editor at *Scientific American* who writes about the intersection of science and society.

MORE TO EXPLORE

Against the Tide: The Battle for America's Beaches. Cornelia Dean. Columbia University Press, 1999.

Taking Chances: The Coast after Hurricane Sandy. Edited by Karen M. O'Neill and Daniel J. Van Abs. Rutgers University Press, 2016.

The Water Will Come: Rising Seas, Sinking Cities, and the Remaking of the Civilized World. Jeff Goodell. Little, Brown, 2017.

Rising: Dispatches from the New American Shore. Elizabeth Rush. Milkweed Editions, 2018.

FROM OUR ARCHIVES

Storm of the Century Every Two Years. Mark Fischetti; June 2013.

scientificamerican.com/magazine/sa

© 2018 Scientific American

EVOLUTION

THE UNLIKELY TRIUMPH OF DINOSAURS

New fossils and analyses topple the long-standing explanation of how dinosaurs came to rule the earth

By Stephen Brusatte

Illustration by James Gurney

IMPROBABLE CHAMPION: *Dromomeron,* a dinosaur precursor, warily approaches the water's edge for a drink 212 million years ago in an oasis in what is now Ghost Ranch in New Mexico. *Koskinonodon,* a giant amphibian, lies in wait.

© 2018 Scientific American

© 2018 Scientific American

WHEN I WAS A TEENAGER AT THE TURN OF THE MILLENNIUM, right around the time I became smitten with fossils, the Field Museum in Chicago dismantled its *Brachiosaurus* and installed a *Tyrannosaurus rex*. In essence, the institution was trading one dinosaur icon for another. Out went the plant-eating colossus, heavier than 10 elephants, its neck arcing gracefully far above the museum's second-floor viewing gallery. In came the biggest, baddest predator of all time: a bus-sized brute with railroad-spike teeth that shattered the bones of its prey.

These were the dinosaurs that fired my imagination as I grew up 75 miles down the road from Chicago, in a flat expanse of Midwestern corn and bean fields. I visited them as often as I could convince my parents to make the drive. Standing underneath their skeletons was hypnotic: their size, their strength, their bodies so alien compared with those of any animals alive today. No wonder they ruled the earth for more than 150 million years. They were magnificent.

But how did dinosaurs get this way? It was a question I rarely contemplated during those obsessive years. In the same way that it was hard to envision that my parents were once my age, I just assumed dinosaurs materialized at some point in the deep past as fully formed long-necked and sharp-toothed giants. I didn't know it at the time, but that wasn't too far off from the scientific consensus for much of the late 20th century. Dinosaurs were special, this convention held, endowed with such superior speed, agility and metabolism that they quickly and easily outcompeted their early rivals, spread across the planet and established an empire.

Over the past 15 years, however, a wealth of new fossil discoveries from around the globe, fresh insights into the physical world the first dinosaurs inhabited, and novel approaches to building family trees and analyzing evolutionary trends have challenged that long-standing view. From these advances, a rather different story has emerged: the rise of dinosaurs was gradual, and for the first 30 million years of their history they were restricted to a few corners of the world, outpaced by other species. Only after catching a couple of lucky breaks did they rise up to take over the planet.

HUMBLE ORIGINS

LIKE MANY SUCCESSFUL ORGANISMS, dinosaurs were born of catastrophe. Around 252 million years ago, at the tail end of the Permian Period, a pool of magma began to rumble underneath Siberia. The animals living at the surface—an exotic menagerie of large amphibians, knobby-skinned reptiles and flesh-eating forerunners of mammals—had no inkling of the carnage to come. Streams of liquid rock snaked through the mantle and then the crust, before flooding out through mile-wide cracks in the earth's surface. For hundreds of thousands, maybe millions, of years the eruptions continued, spewing heat, dust, noxious gases and enough lava to drown several million square miles of the Asian landscape. Temperatures spiked, oceans acidified, ecosystems collapsed and up to 95 percent of the Permian species went extinct. It was the worst mass extinction in our planet's history. But a handful of survivors staggered into the next period of geologic time, the Triassic. As the volcanoes quieted and ecosystems stabilized, these plucky creatures now found themselves in a largely empty world. Among them were various small amphibians and reptiles, which diversified as the earth healed and which later diverged into today's frogs, salamanders, turtles, lizards and mammals.

Scientists know about these animals from the handprints and footprints they left in layer after layer of river and lake sediments now exposed in the Holy Cross Mountains in Poland. For more than 20 years Grzegorz Niedźwiedzki, who grew up in these hills and is now a paleontologist based at Uppsala University in Sweden, has meticulously collected these fossil tracks, occasionally with me by his side. In 2005, while fossil hunting

IN BRIEF

The conventional view of dinosaur origins holds that they were endowed with such superior speed, agility, metabolism and intelligence that as soon as they debuted they quickly outcompeted the competition and spread across the globe.

New fossil discoveries and analyses have upended that scenario, showing that dinosaurs languished on the evolutionary sidelines for tens of millions of years before they were able to surpass their rivals and conquer the planet.

© 2018 Scientific American

GRZEGORZ NIEDŹWIEDZKI *Uppsala University*

FOSSIL TRACKS of *Prorotodactylus* **show that around 250 million years ago dinosaur precursors called dinosauromorphs roamed what are now the Holy Cross Mountains in Poland.**

near the village of Stryczowice, along a narrow stream tangled with brambles, Niedźwiedzki discovered an unusual type of track that did not seem to match any of the more common reptile and amphibian traces. The strange prints are about the size of a cat's paw, arranged in narrow trackways, with the five-fingered handprints positioned in front of the slightly larger footprints, which have three long central toes flanked by a toe nubbin on each side. The tracks go by the genus name *Prorotodactylus*. All that we know about this creature comes from these prints—there are no known fossils of the animal itself.

The *Prorotodactylus* tracks date to about 250 million years ago, just one or two million years after the volcanic eruptions that brought the Permian to a close. Early on it was clear from the narrow distance between the left and right tracks that they belonged to a specialized group of reptiles called archosaurs that emerged after the Permian extinction with a newly evolved upright posture that helped them run faster, cover longer distances and track down prey with greater ease. The fact that the tracks came from an early archosaur meant that they could potentially bear on questions about the origins of dinosaurs. Almost as soon as the archosaurs originated, they branched into two major lineages, which would grapple with each other in an evolutionary arms race over the remainder of the Triassic: the pseudosuchians, which led to today's crocodiles, and the avemetatarsalians, which developed into dinosaurs. Which branch did *Prorotodactylus* belong to?

To find out, I conducted a study with Niedźwiedzki and Richard J. Butler, now at the University of Birmingham in England. Our analysis of the prints, published in 2011, revealed peculiarities of the footprints that link them to signature features of the dinosaur foot: the digitigrade arrangement of the bones, in which only the toes make contact with the ground while walking, and the very narrow foot with three main toes. *Prorotodactylus* is therefore a dinosauromorph: not a dinosaur per se but a primitive member of the avemetatarsalian subgroup that includes dinosaurs and their very closest cousins. Members of this group had long tails, big leg muscles, and hips with extra bones connecting the legs to the trunk, which allowed them to move even faster and more efficiently than other archosaurs.

These earliest dinosauromorphs were hardly fearsome, however. Fossils indicate that they were only about the size of a house cat, with long, skinny legs. And there were not very many of them either: less than 5 percent of all Stryczowice tracks belong to *Prorotodactylus,* which is far outnumbered by tracks of small reptiles, amphibians and even other archosaurs. The dinosauromorphs' time had not come. Yet.

THE FIRST DINOSAURS

OVER THE NEXT 10 MILLION to 15 million years the dinosauromorphs continued to diversify. The fossil record from this time period shows an increasing number of track types in Poland and then around the world. The tracks get larger and develop a greater variety of shapes. Some trackways stop showing impressions of the hand, a sign the makers were walking only on their hind legs. Skeletons start to turn up as well. Then, at some point between 240 million and 230 million years ago, one of these primitive dinosauromorph lineages evolved into true dinosaurs. It was a radical change in name only—the transition involved just a few subtle anatomical innovations: a long scar on the

© 2018 Scientific American

upper arm that anchored bigger muscles, some tablike flanges on the neck vertebrae that supported stronger ligaments, and an open, windowlike joint where the thighbone meets the pelvis that stabilized upright posture. Still, modest though these changes were, they marked the start of something big.

The oldest unequivocal dinosaur fossils, which date to around 230 million years ago, come from the otherworldly landscapes of Ischigualasto Provincial Park in Argentina. Scientists have collected there for decades, beginning with legendary American paleontologist Alfred Romer in the 1950s and continuing with Argentine researchers Osvaldo Reig and José Bonaparte in the 1960s. More recently, Paul Sereno of the University of Chicago and Ricardo N. Martínez of the National University of San Juan in Argentina led expeditions to Ischigualasto in the 1980s and 1990s. Among the fossils they found there were those belonging to *Herrerasaurus, Eoraptor* and other creatures representing all three of the main branches of the dinosaur family: the meat-eating theropods; the long-necked, plant-eating sauropodomorphs and the beaked, plant-eating ornithischians.

By the middle part of the Triassic, around 230 million to 220 million years ago, these three main dinosaur subgroups were on the march, siblings setting out to form their own broods in a world we would barely recognize. Back then a single supercontinent called Pangea stretched from pole to pole, surrounded by a global ocean called Panthalassa. It was not a safe place to call home. The earth was much warmer, and because Pangea was centered on the equator, half the land was always scorching in the summer while the other half was cooler in the winter. These marked temperature differences fueled violent "mega monsoons" that divided Pangea into environmental provinces characterized by varying degrees of precipitation and wind. The equatorial region was unbearably hot and muggy, flanked by subtropical deserts on both sides. The midlatitude regions were slightly cooler and much wetter.

Herrerasaurus, Eoraptor and the other Ischigualasto dinosaurs were ensconced in the comparatively hospitable midlatitudes. So were their counterparts from Brazil and India, known from exciting recent fossil discoveries. But what about other parts of the supercontinent? Did early dinosaurs colonize these harsher regions just as capably, as the conventional wisdom about them suggests? In 2009, a few months after our first jaunt together in Poland, Butler and I teamed up with Octávio Mateus of the Museum of Lourinhã in Portugal to test this hypothesis by exploring a remnant of the subtropical arid belt of northern Pangea in what is now southern Portugal. We were hoping to find dinosaurs, but what we found instead was a mass graveyard of hundreds of Smart car–sized amphibians that we assigned to a new species, *Metoposaurus algarvensis*. These rulers of the Triassic lakes and rivers had been victims of a freak shift in the capricious Pangean weather that probably caused their lakes to dry up. We returned later to excavate the bone bed and started to also find fossils of various fishes, poodle-sized reptiles and archosaurs from the line leading to crocodiles. But still, to this day, we have yet to come across even a scrap of dinosaur bone.

We probably never will. Spain, Morocco and the eastern seaboard of North America have stellar fossil sites from this same time between 230 million and 220 million years ago that show the same pattern we saw in Portugal: plenty of amphibians and reptiles but nary a dinosaur. All these places were in the arid sec-

Family Feud

Perhaps the most heated debate in contemporary dinosaur research concerns how the theropods, sauropodomorphs and ornithischians are arranged on the family tree. In 1887 British paleontologist Harry Govier Seeley surveyed the flood of new fossils from Europe and the American West and argued that dinosaurs could be separated into two distinct types, based on the structure of their hip bones. Theropods and sauropodomorphs both have a pubis bone pointing forward, as modern lizards do, so he placed them together in a group he called Saurischia—the "lizard-hipped" species. Ornithischians, with their pubis projecting backward like that of modern birds, were deemed a separate branch of "bird-hipped" dinosaurs. This basic dichotomy persists today as the standard dinosaur classification scheme that I and all my fellow dinosaur hunters learned as students.

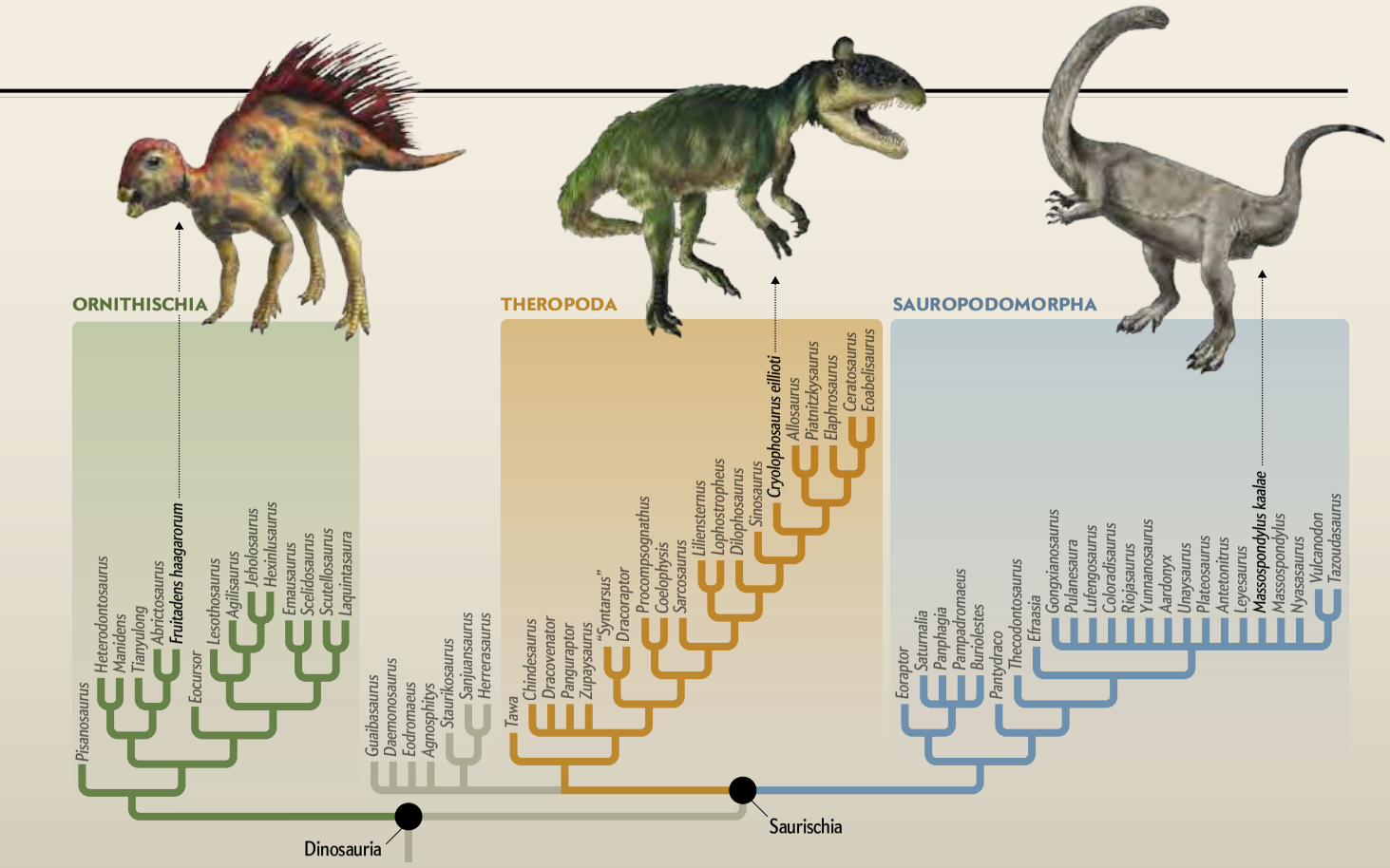
It might be incorrect, however. In a bombshell study published in *Nature* early last year, University of Cambridge Ph.D. student Matthew Baron and his colleagues presented a new dinosaur genealogy based on an analysis of an expansive data set of early dinosaurs and their anatomical features. Their tree links together theropods and ornithischians into a group they call Ornithoscelida, with sauropodomorphs perched outside on a separate limb. Instead of saurischians versus ornithischians, the new dinosaur dichotomy is ornithoscelidans versus sauropodomorphs.

Or maybe not. Soon after Baron's study was released, I was approached by Max C. Langer, a Brazilian paleontologist who has described a slew of new Triassic dinosauromorphs and dinosaurs from his homeland over the past decade, including *Ixalerpeton* (a dinosaur precursor very similar to the one that left the *Prorotodactylus* tracks from Poland) and *Saturnalia* (a dog-sized protosauropodomorph). He was skeptical of the new genealogy and recruited a team of experts on early dinosaurs to pore over Baron's data set. Because I had studied the Polish trackways and other key Triassic fossils, Langer asked me to be part of the group. For a month we carefully picked through the data set and noted our various disagreements about how the other team had characterized certain features. We then reran the analysis of the traits with our corrections. The resulting family tree shifted back to saurischians versus ornithischians, although statistical tests showed that this arrangement was not a significantly better fit to the data than Baron's ornithoscelidan versus sauropodomorph tree. We presented our results in a follow-up *Nature* paper in the autumn of 2017.

What this ambivalence in the results means is that paleontologists do not currently have a good understanding of the basic shape of the dinosaur tree. It seems that the rush of new discoveries in Argentina, Brazil, Poland and elsewhere over the past 15 years has muddied the picture. We now realize that the earliest members of the three major dinosaur lineages were remarkably similar in body size and anatomy, which makes untangling their relationships difficult. This puzzle is ripe to be solved by the next generation of paleontologists, probably the way these arguments are usually settled: with new fossils. —*S.B.*

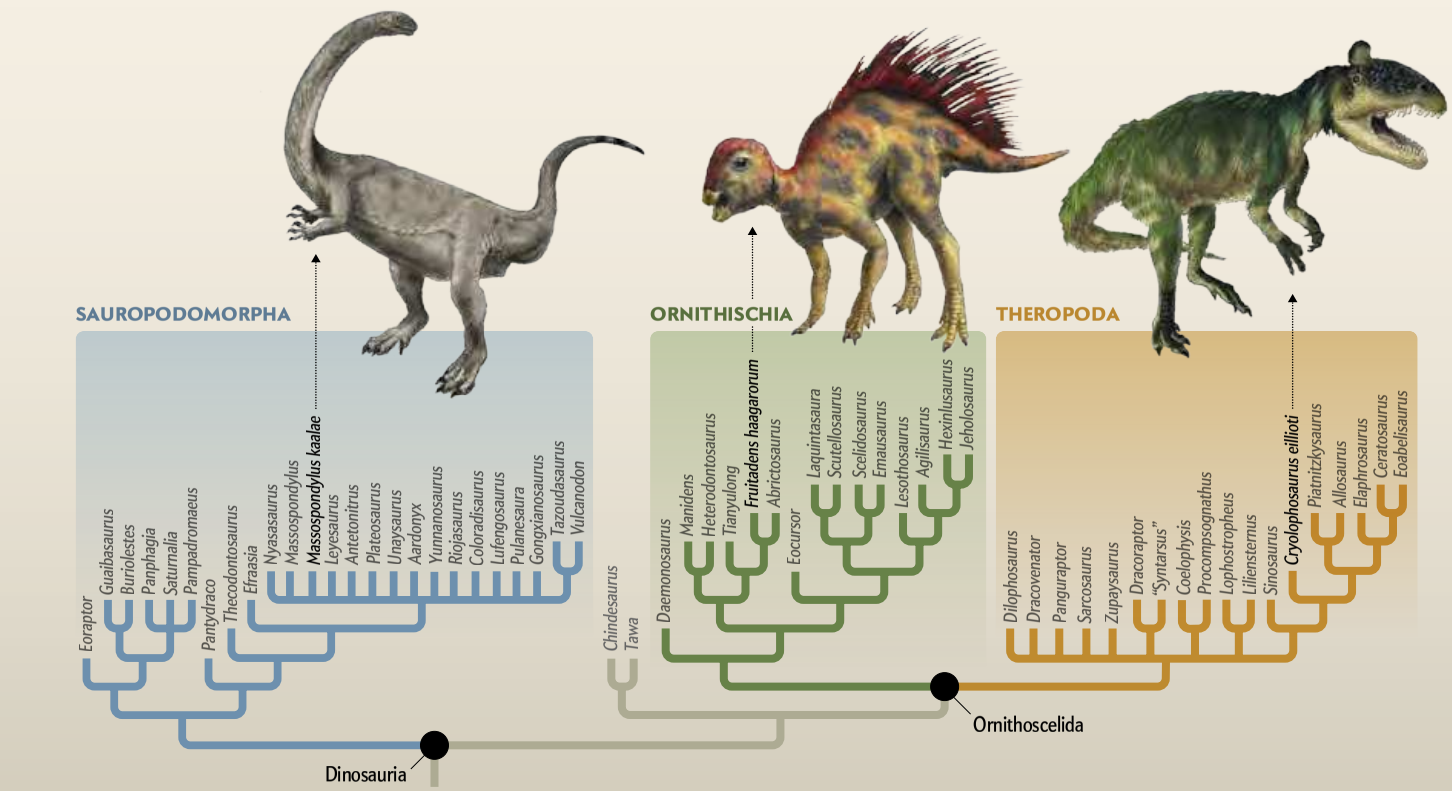
SOURCES: "UNTANGLING THE DINOSAUR FAMILY TREE." BY MAX C. LANGER ET AL., IN *NATURE*, VOL. 551; NOVEMBER 2, 2017; "BARON ET AL. REPLY." BY MATTHEW G. BARON ET AL., IN *NATURE*, VOL. 551. PUBLISHED ONLINE NOVEMBER 1, 2017

© 2018 Scientific American



Traditional View

Paleontologists have long divided dinosaurs into the bird-hipped Ornithischia and the lizard-hipped Saurischia, which includes the theropods and sauropodomorphs.



New Hypothesis

A recent analysis of dinosaur traits concluded that theropods and ornithischians belong to a group called Ornithoscelida and that sauropodomorphs sit on a separate branch.

Illustrations by Portia Sloan Rollings, Graphics by Jen Christiansen

© 2018 Scientific American

tor of Pangea. Together these sites indicate that during the formative years of their evolution, dinosaurs were slowly diversifying in the humid temperate regions but were seemingly unable to colonize the deserts. It is an unexpected story line: far from being superior creatures that swept across Pangea the moment they originated, dinosaurs could not handle the heat. They were geographically localized—mere bit players in the drama playing out across a world still recovering from the great End Permian extinction.

But then, just when it seemed that dinosaurs would never escape their rut, they received two lucky breaks. First, in the humid zone, the dominant large herbivores of the time—reptiles called rhynchosaurs and mammal cousins called dicynodonts—went into decline, disappearing entirely in some areas for reasons still unknown. Their fall from grace between 225 million and 215 million years ago gave primitive plant-eating sauropodomorphs such as Saturnalia, a dog-size species with a slightly elongated neck, the opportunity to claim an important niche. Before long these sauropod precursors were the main herbivores in the humid parts of the Northern and Southern Hemispheres. Second, around 215 million years ago dinosaurs finally broke into the deserts of the Northern Hemisphere, probably because shifts in the monsoons and the amount of carbon dioxide in the atmosphere made differences between the humid and arid regions less severe, allowing dinosaurs to migrate between them more easily.

They still had a long road ahead of them, however. The best records of these first desert-dwelling dinosaurs come from areas that are once again deserts today, in the colorful badlands of the southwestern U.S. For more than a decade a team of young researchers has been methodically excavating the Hayden Quarry, a fossil-rich locality in artist Georgia O'Keeffe's much loved retreat of Ghost Ranch in New Mexico. Randall Irmis of the University of Utah, Sterling Nesbitt of Virginia Tech, Nathan Smith of the Natural History Museum of Los Angeles County, Alan Turner of Stony Brook University and Jessica Whiteside of the University of Southampton in England have found a bounty of skeletons: monster amphibians closely related to our Portuguese *Metoposaurus,* primitive crocodile relatives, and a host of curious swimming and tree-hopping reptiles. There are also dinosaurs in the Hayden Quarry, though not many of them: only a few species of predatory theropods, each represented by a few fossils. There were no plant eaters: none of the ancestral long-necked species so common in the humid zones, none of the ornithischian forebears of *Triceratops.* The team argued that, once again, the paucity of dinosaurs came down to the weather: these deserts were unstable environments of fluctuating temperatures and rainfall, with raging wildfires during some parts of the year and humid spells in others. Plants had difficulty establishing stable communities, which meant that plant-eating dinosaurs did not have a steady source of food. Thus, some 20 million years after they had originated and even after they had taken over the big herbivore role in humid ecosystems and started to settle the tropical deserts, dinosaurs had yet to mount a global revolution.

STIFF COMPETITION: For much of the Triassic period dinosaurs were a marginal group, overshadowed by the likes of crocodile relatives such as *Saurosuchus* (*1*) and giant amphibians such as *Metoposaurus* (*2*).

RICARDO N. MARTÍNEZ *Institute and Museum of Natural Sciences, National University of San Juan* (1); TOMASZ SULEJ *Institute of Paleobiology, Polish Academy of Sciences* (2)

CROC COMPETITION

NO MATTER WHICH INTERVAL you look at in the Triassic, from the time the first dinosaurs appeared around 230 million years ago until the period ended 201 million years ago, the story is the same. Only some dinosaurs were able to live in some parts of the world, and wherever they lived—humid forests or parched deserts—they were surrounded by all kinds of bigger, more common, more diverse animals. In Argentina's Ischigualasto, for instance, those earliest dinosaurs made up only about 10 to

© 2018 Scientific American

20 percent of the total ecosystem. The situation was similar in Brazil and, millions of years later, at the Hayden Quarry. In all cases, the dinosaurs were vastly outnumbered by mammal forerunners, giant amphibians and eccentric reptiles.

More than anything, however, Triassic dinosaurs were being outgunned by their close cousins the so-called pseudosuchians, on the crocodile side of the archosaur family. At Ischigualasto, a crocodile-line archosaur called *Saurosuchus* ruled the food chain, with its sharp teeth and muscular jaws. Hayden Quarry harbored numerous pseudosuchian species: semiaquatic ones with long snouts, armored ones that ate plants, and even toothless ones that sprinted on their hind legs and bore a striking resemblance to some of the theropod dinosaurs they lived alongside.

As a master's student in the late 2000s, around the time many of these fossils were being discovered, I found this pattern peculiar. At the same time I was following the onslaught of new fossils, I started reading classic studies by giants in the field of paleontology, including Robert Bakker and Alan Charig, who effusively argued that dinosaurs were so perfectly adapted, with speed and endurance and smarts, that they quickly took out their crocodile cousins and other competitors during the Triassic. But this idea did not seem to jibe with the fossil record. Was there some way I could test it?

After immersing myself in literature on statistics, I realized that two decades earlier paleontologists who study invertebrate animals had come up with a method for measuring anatomical diversity in a group of species, which had so far been ignored by dinosaur researchers. This measurement is called morphological disparity. If I could track the disparity of dinosaurs and pseudosuchians over the Triassic, I could see whether they were becoming more or less diverse and at what rate—which would indicate whether they became successful gradually or abruptly—and whether one group was pulling ahead of the other.

Working with my then supervisors at the University of Bristol in England—Michael Benton, Marcello Ruta and Graeme Lloyd—I compiled a large data set of Triassic dinosaurs and pseudosuchians, assessing more than 400 characteristics of their anatomy. When we analyzed it statistically, we came up with a startling result that we published in 2008 in *Science*. All throughout the Triassic the pseudosuchians were significantly more anatomically diverse than the dinosaurs, which indicates that they were experimenting with more diets, more behaviors and more ways of making a living. Both groups were becoming more diverse as the Triassic unfolded, but the pseudosuchians always outpaced the dinosaurs. Contrary to the leading view of dinosaurs as superior soldiers slaying their rivals, they were actually losing to the pseudosuchians for most of their long coexistence.

CARPE DIEM

OUR STATISTICAL ANALYSIS led us to an iconoclastic conclusion: the first dinosaurs were not particularly special, at least compared with the variety of other animals they were evolving alongside during the Triassic. If you were around back then to survey the Pangean scene, you probably would have considered the dinosaurs a fairly marginal group. And if you were of a gambling persuasion, you would probably have bet on some of the other animals, most likely those hyperdiverse pseudosuchians, to eventually become dominant, grow to massive sizes and conquer the world. But of course, we know that it was the dinosaurs that became ascendant and even persist today as more than 10,000 species of birds. In contrast, only two dozen or so species of modern crocodilians have survived to the present day.

How did dinosaurs eventually wrestle the crown from their crocodile-line cousins? The biggest factor appears to have been another stroke of good fortune outside the dinosaurs' control. Toward the end of the Triassic great geologic forces pulled on Pangea from both the east and west, causing the supercontinent to fracture. Today the Atlantic Ocean fills that gap, but back then it was a conduit for magma. For more than half a million years tsunamis of lava flooded across much of central Pangea, eerily similar to the enormous volcanic eruptions that closed out the Permian 50 million years prior. Like those earlier eruptions, the End Triassic ones also triggered a mass extinction. The crocodile-line archosaurs were decimated, with only a few species—the ancestors of today's crocodiles and alligators—able to endure.

Dinosaurs, on the other hand, seemed to have barely noticed this fire and brimstone. All the major subgroups—the theropods, sauropodomorphs and ornithischians—sailed into the next interval of geologic time, the Jurassic Period. As the world was going to hell, dinosaurs were thriving, somehow taking advantage of the chaos around them. I wish I had a good answer for why—was there something special about dinosaurs that gave them an edge over the pseudosuchians, or did they simply walk away from the plane crash unscathed, saved by sheer luck when so many others perished? This is a riddle for the next generation of paleontologists to solve.

Whatever the reason dinosaurs survived that disaster, there is no mistaking the consequences. Once on the other side, freed from the yoke of their pseudosuchian rivals, these dinosaurs had the opportunity to prosper in the Jurassic. They became more diverse, more abundant and bigger than ever before. Completely new dinosaur species evolved and migrated widely, taking pride of place in terrestrial ecosystems the world over. Among these newcomers were the first dinosaurs with plates on their backs and armor covering their bodies; the first truly colossal sauropods that shook the earth as they walked; carnivorous ancestors of *T. rex* that began to get much bigger; and an assortment of other theropods that started to get smaller, lengthen their arms and cover themselves in feathers—predecessors of birds. Dinosaurs were now dominant. It took more than 30 million years, but they had, at long last, arrived. SA

Stephen Brusatte is a paleontologist at the University of Edinburgh in Scotland. His research focuses on the anatomy and evolution of dinosaurs. He is author of the new book *The Rise and Fall of the Dinosaurs* (William Morrow, 2018).

MORE TO EXPLORE

The Origin and Early Radiation of Dinosaurs. Stephen L. Brusatte et al. in *Earth-Science Reviews,* Vol. 101, Nos. 1–2, pages 68–100; July 2010.

Footprints Pull Origin and Diversification of Dinosaur Stem Lineage Deep into Early Triassic. Stephen L. Brusatte, Grzegorz Niedźwiedzki and Richard J. Butler in *Proceedings of the Royal Society B,* Vol. 278, pages 1107–1113; April 7, 2011.

Untangling the Dinosaur Family Tree. Max C. Langer et al. in *Nature,* Vol. 551, pages E1–E3; November 2, 2017.

FROM OUR ARCHIVES

Rise of the Tyrannosaurs. Stephen Brusatte; May 2015.

scientificamerican.com/magazine/sa

© 2018 Scientific American

GEOLOGY

Canary in the Crater

Scientists scramble to analyze data from the eruption of Kilauea, hoping to predict similar events elsewhere

By Sara Reardon

MARIO TAMA *Getty Images*

After weeks of unleashing earthquakes and lava flows that have forced thousands of people to evacuate their homes, Hawaii's Kilauea volcano has finally blown its top. Because Kilauea is one of the best-monitored volcanoes in the world, scientists hope that data on the event will help them to better predict when similar volcanoes are about to erupt. "We'll be working on this set of data for our careers," says Michael Poland, a geophysicist at the U.S. Geological Survey (USGS) Cascades Volcano Observatory in Vancouver, Wash.

The USGS says that the eruption began at 4:15 A.M. local time on May 17, when the volcano sent a plume of ash and steam more than 9,100 meters into the air.

The many instruments that had been placed on and around Kilauea were watching. The volcano bristles with equipment that continuously measures signs of geologic activity, such as ground movement, lava chemistry and seismic vibrations.

The first hint of an impending eruption came with a series of earthquakes on May 3. Soon after, fissures opened up in the ground as far as 40 kilometers away from the volcano's rim—oozing lava that forced about 2,000 people to evacuate. The openings also depressurized the network of underground channels underneath Kilauea, including its lava chamber. As a result, the lava level within the volcano's crater quickly dropped by more than 30 meters. It was, Poland says, "like someone pulled the plug in a bathtub."

That caused the walls of the volcano to begin crumbling into the crater, creating a layer of rock atop the surface of the remaining lava. And once the surface of the still draining lava dropped below the water table, water began to seep into the crater, generating steam and pressure below the freshly formed rock cap.

© 2018 Scientific American

ASH PLUME rises in the distance from the Kilauea volcano on Hawaii's Big Island on May 15, 2018.

HISTORY REPEATS

SCIENTISTS at the USGS's nearby Hawaiian Volcano Observatory suspected that a steam explosion was imminent: in 1924 the same pattern of oozing fissures around Kilauea had heralded a series of explosions.

The researchers were proved right when the pressure that built up in the crater sent debris and ash flying. But Poland says that steam explosions are hard to anticipate.

The hope this time, he adds, is that the extensive instrumentation on Kilauea and data collected from the latest eruption will allow scientists to develop better markers for predicting when a steam explosion is imminent.

Such events are "very small and localized, but if you're in the neighborhood it's going to cause an impact," Poland says. In 2014 debris and ash from a steam explosion at Mount Ontake in Japan killed 63 people.

For now scientists in Hawaii and around the world are watching Kilauea and waiting to see what else the volcano will do. "This has so far been playing out how USGS said it would," says Janine Krippner, a volcanologist at Concord University in Athens, W.V. "That itself is incredible."

Kilauea probably isn't done erupting yet, if the 1924 event is a guide: it went on for more than two weeks and caused more than 50 steam explosions. Hawaii's state volcanologist, Bruce Houghton, says that the current event seems to be a type of complex eruption that occurs only once every 50 to 80 years.

SAFETY FIRST

THE RECENT ACTIVITY has prompted the state of Hawaii to order the evacuation of several areas near Kilauea, protecting residents from the greatest eruption danger—flying rocks. But ash could still present a problem for people with respiratory problems, and officials are monitoring the wind and weather patterns to predict where the ash from the volcano will go.

Houghton says that it may be hard to know when it is safe for people to return to evacuated areas. In the 1960s an eruption at Kilauea stopped for a year and a half before the volcano suddenly resumed spewing ash. "It takes quite a long time to be certain things are completely over," he says. "It will suit us just fine if it stops tomorrow, but the past record suggests it might go for months or years."

Kilauea's latest eruption "is an exciting event, but it's coming at a cost," Poland says, noting that dozens of homes have been destroyed and tourism is suffering. "We as the scientific community feel we owe it to the people being impacted to get it right and learn as much as we can," he says. SA

Editors' note: *Since this article was published, Kilauea continued to erupt and trigger minor earthquakes until August 2, 2018, when things quieted down. As of this printing, the volcanic activity seems to have paused, although it is not necessarily over. About 700 homes on the Big Island were destroyed by flowing lava. No one was killed.*

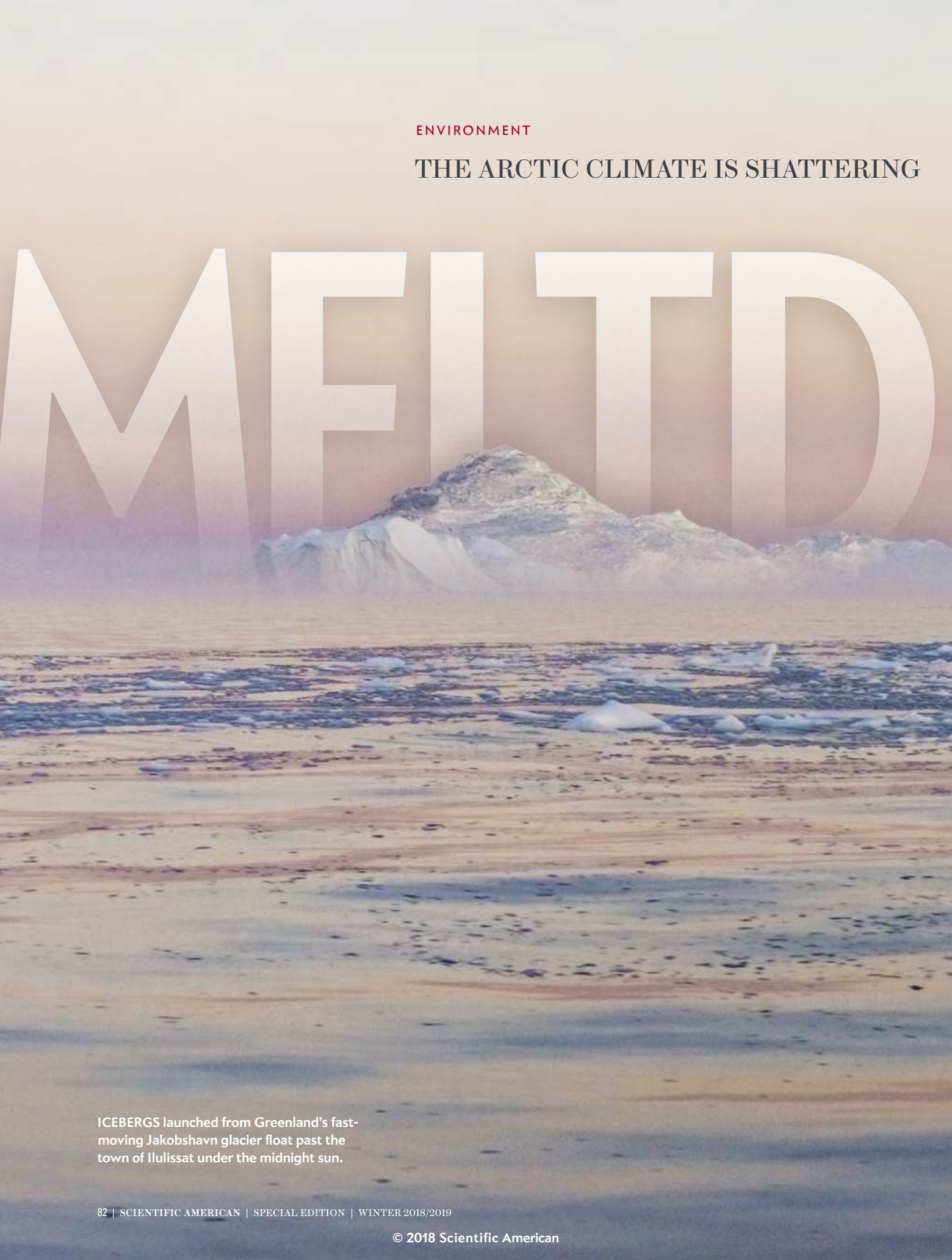
Sara Reardon is a senior reporter at *Nature* in Washington, D.C., covering biomedical research and policy. She has previously written for *New Scientist* and *Science* and has a master's degree in molecular biology from the University of Washington.

© 2018 Scientific American

ENVIRONMENT

THE ARCTIC CLIMATE IS SHATTERING

MELTD



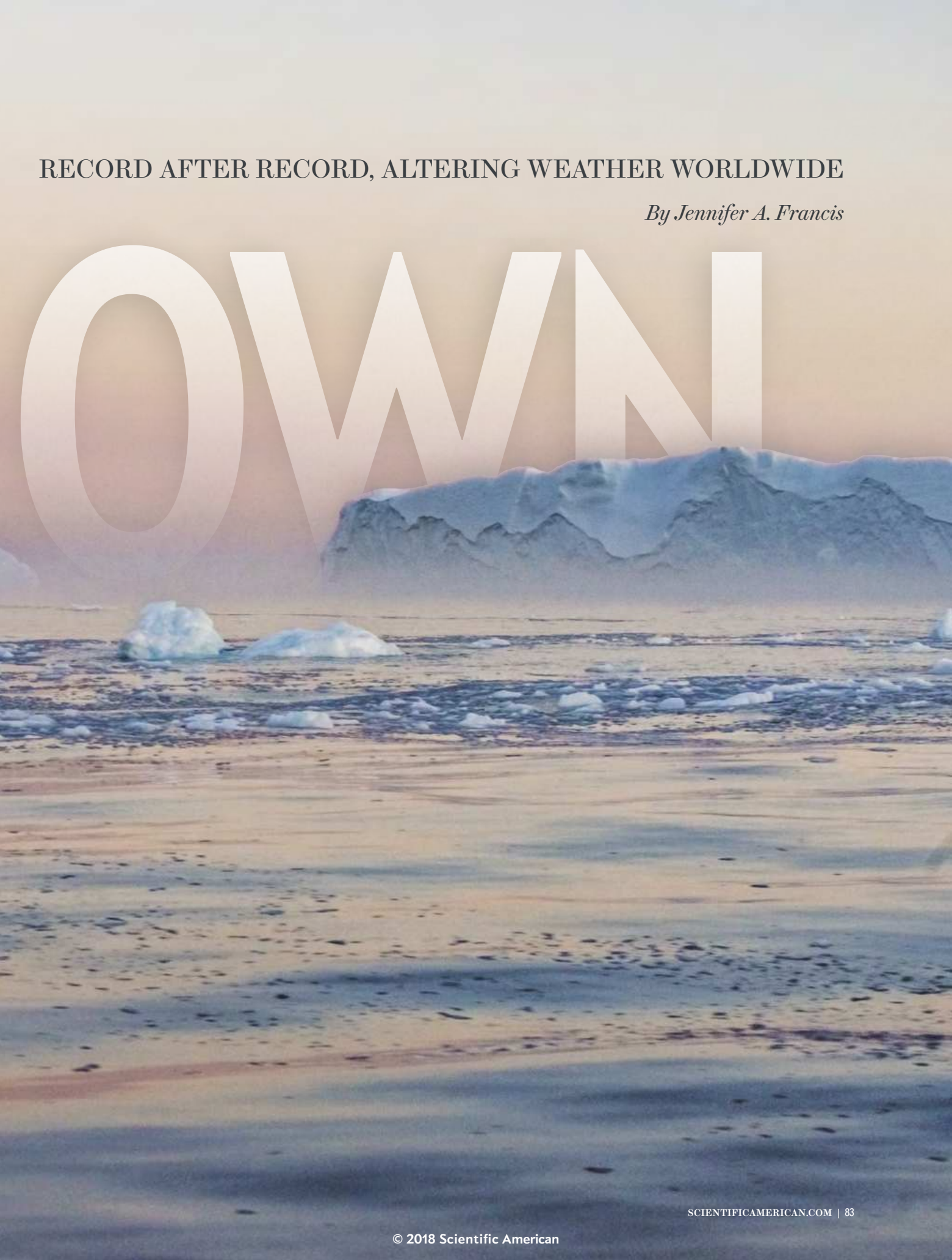
ICEBERGS launched from Greenland's fast-moving Jakobshavn glacier float past the town of Ilulissat under the midnight sun.

© 2018 Scientific American

OWN

RECORD AFTER RECORD, ALTERING WEATHER WORLDWIDE

By Jennifer A. Francis



© 2018 Scientific American

TWENTY-FIVE SCIENTISTS, including me, had an epiphany about the Arctic in 2003. The National Science Foundation had invited us to a retreat in Big Sky, Mont. Before this gathering, each of us had been focusing our Arctic research on our own narrow topics. As we shared our perspectives, we came to a frightening realization: the changes we had been finding individually were connected. They fit together perfectly. The Arctic system as a whole was careening toward a precarious new state. And hope of stopping it already seemed unlikely.

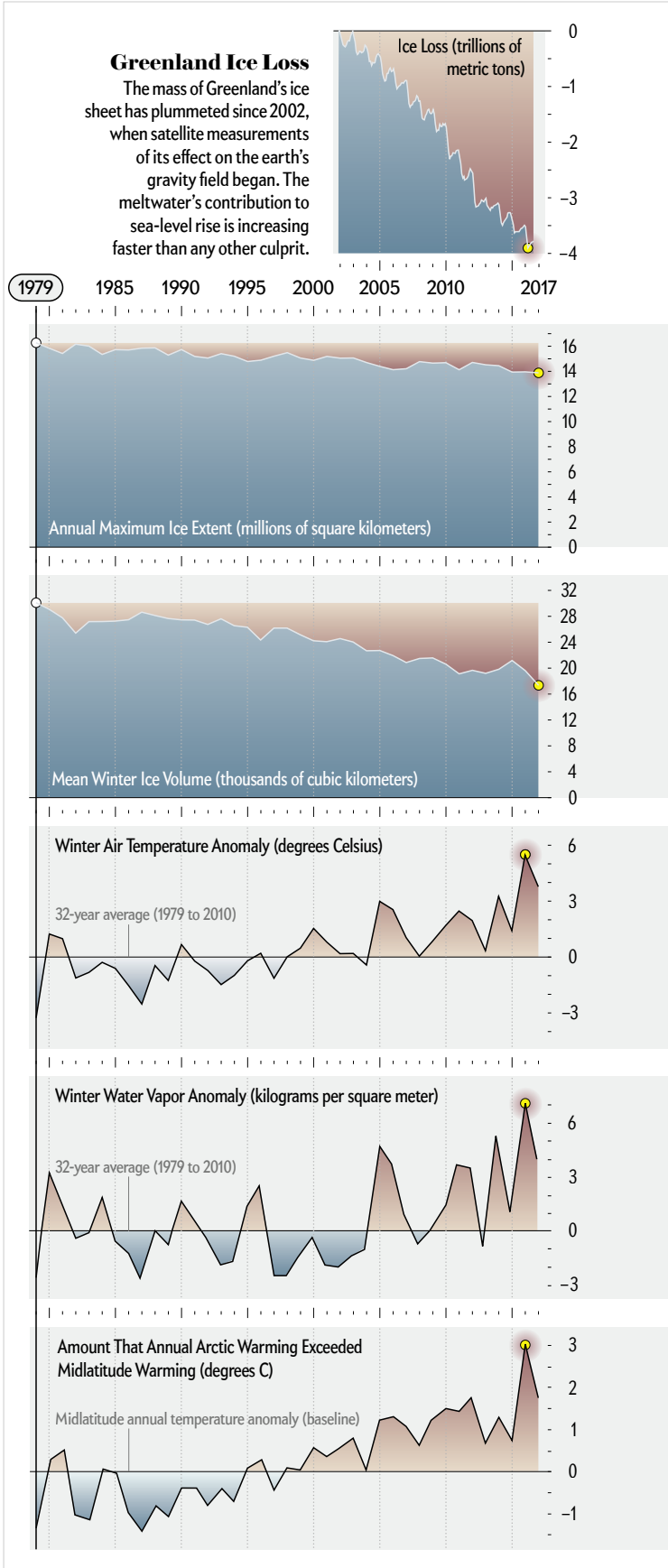
We published a paper with a stunning, controversial conclusion: At the current rate of change, there was a real possibility that within a century, the world could witness a summer Arctic Ocean that would be ice-free, a state not seen for thousands of years. Today I am startled again because it now appears that the ocean will likely be free of summer ice by 2040—a full 60 years earlier than we had predicted little more than a decade ago.

The Arctic is changing exactly the way scientists thought it would but faster than even the most aggressive predictions. The recent behavior is off the charts. In just three years, more than a dozen climate records that had each stood for many decades have crumbled, including those for disappearing summer sea ice, decreasing winter sea ice, warming air and thawing ground.

These trends signal trouble for people around the world. The last time the Arctic was only slightly warmer than today—about 125,000 years ago—oceans were 13 to 20 feet higher. Goodbye Miami, New Orleans, the naval base in Norfolk, Va., most of New York City and Silicon Valley, as well as Venice, London and Shanghai. New research suggests that rapid Arctic warming also tends to reroute the jet stream in ways that could allow punishing weather patterns to linger across North America, central Europe and Asia longer than usual, subjecting millions of people to unyielding heat waves, droughts or relentless storms. Plankton are increasing throughout the southern Arc-

IN BRIEF

The Arctic climate is changing rapidly, breaking at least a dozen major records in the past three years. **Sea ice** is disappearing, air temperatures are soaring, permafrost is thawing and glaciers are melting. **The swift warming is altering** the jet stream and polar vortex, prolonging heat waves, droughts, deep freezes and heavy rains worldwide.



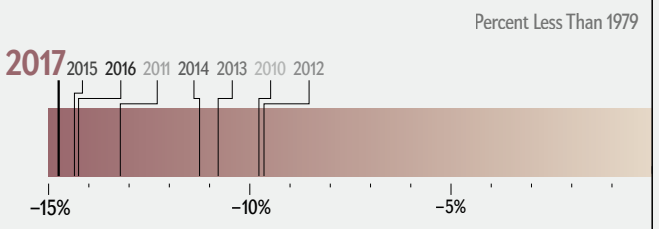
PRECEDING PAGES: GETTY IMAGES

© 2018 Scientific American

Record-Breaking Arctic

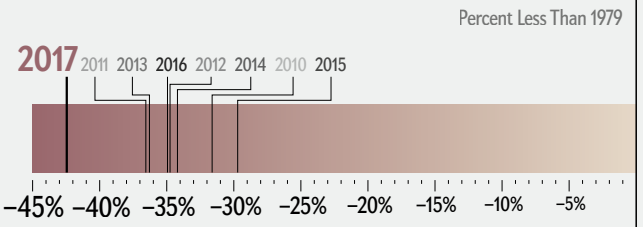
The Arctic region is changing dramatically, and the rapid transformation will affect billions of people around the world. In the past three years alone numerous climate records up north have been broken, in some cases by stunning margins. Six notable examples are shown here. Red numbers in the bar charts indicate all-time records; values for recent years reflect a trend toward those excesses. Individually, the effects alter the environment and daily life for people throughout the region. Together the effects are reshaping weather across the Northern Hemisphere, culminating in Arctic amplification (*final graph*), which raises the chance for extreme conditions year-round.

Value in 1979



Winter Sea Ice Extent

As winter deepens, ice spreads over the Arctic Ocean. But the maximum reach has declined steadily, especially across the Barents and Bering seas. Less ice cover allows the open ocean to send more heat and moisture into the air, significantly disrupting Arctic weather.

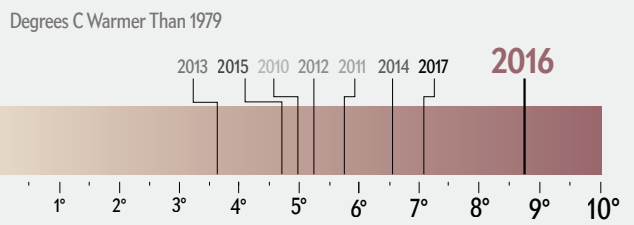


Winter Sea Ice Volume

By 2017 the amount of winter ice floating on the Arctic Ocean had dropped an astounding 42.5 percent since 1979. Winds can more easily push thin ice, trapping ships and coastal communities. Thinner ice also melts faster in warm months; summer ice volume has dropped 80 percent in the same time span.

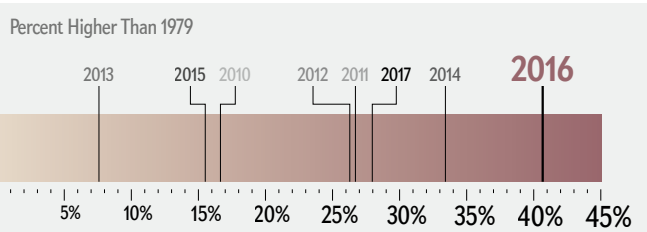
Winter Air Temperature

On certain days Arctic temperatures can soar 20 degrees Celsius above normal, and they are now elevated throughout the winter. In 2016 the mean winter temperature was almost nine degrees higher than in 1979. This trend can weaken the jet stream, bringing deep cold snaps and snows to the U.S., Europe and Asia.



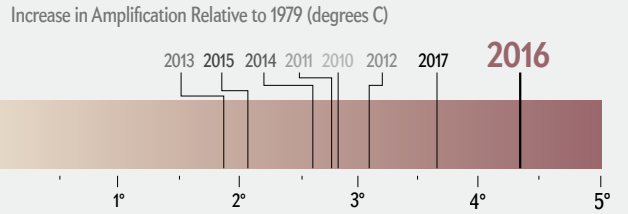
Winter Water Vapor

With less ice cover, more open ocean sends added moisture into the air. Even a small rise has big, underappreciated consequences: water vapor is a greenhouse gas that traps heat. It also condenses into clouds, releasing its latent heat. The clouds can enhance warming, too.



Arctic Amplification

The Arctic is warming faster than the rest of the world. This "amplification" means the average temperature in the Arctic is getting closer to the average temperature in the midlatitudes. The narrowing gap slows the jet stream year-round, raising the chances across the Northern Hemisphere for more persistent, extreme weather patterns—among them heat waves, floods, cold spells and maybe even longer-lasting hurricanes.



SOURCES: JPL GRACE MASCON OCEAN, ICE, AND HYDROLOGY EQUIVALENT WATER HEIGHT RL05M.1 CRI FILTERED VERSION 2. D. N. WIESE, D-N. YUAN, C. BOENING, F. W. LANDERER AND M. M. WATKINS. PO.DAAC, JPL, APRIL 29, 2016. DATA ACCESSED FEBRUARY 14, 2018 (*Greenland ice*); JAPAN'S NATIONAL INSTITUTE OF POLAR RESEARCH AND JAXA. VIA ARCTIC DATA ARCHIVE SYSTEM (*ice extent*); PIOMAS, POLAR SCIENCE CENTER (*ice volume*); NCEP/NCAR REANALYSIS 1, PHYSICAL SCIENCES DIVISION, ESRL, NOAA (*air temperature, water vapor anomaly and amplification data*)

Graphic by Pitch Interactive

© 2018 Scientific American

tic Ocean, which may disrupt food chains that support commercial fisheries. And the massive ice melt is adding to an enormous blob of freshwater south of Greenland that may be slowing the Gulf Stream, which could significantly change weather patterns for continents on both sides of the Atlantic Ocean.

What's driving this breakneck change?

LOSS OF "PERMANENT" ICE

SCIENTISTS EXPEND A LOT OF ENERGY watching the Arctic because it is so sensitive to climatic change. It is a "canary in the coal mine" for the earth's entire climate system. The long list of records that have been smashed during the past several years leaves no doubt that decades of unnerving climate simulations are on target. But the data say something even greater: our projections of changes on the way may be far too conservative.

In only 40 years the extent of ice across the Arctic Ocean in summer has shrunk by half. Yes, half. The volume of sea ice, year-round, is way down, too—about a quarter of what it was in the early 1980s. Until recently, scientists had thought it would take until at least the middle of this century to reach these extremes.

Summer sea ice is vanishing quickly because of feedbacks—vicious cycles that can amplify a small change. For example, when a bit of extra heat melts bright-white ice, more of the dark ocean surface is exposed, which reflects less of the sun's energy back to space. That absorbed heat then warms the area further, which melts even more ice, leading to yet more warming. In winter when the sun is not shining, other feedbacks take over. For example, sea ice floating on the Arctic Ocean acts like an insulating sheet, preventing heat and moisture under it from escaping into the atmosphere. When that ice recedes, more heat and moisture can warm the air, further retarding ice formation. Computer simulations generally show ice disappearing too slowly, contributing to the conservative estimates of future warming.

Disappearing sea ice is not the only transformation keeping researchers like me awake at night. The other two types of what used to be called permanent ice in the Arctic are also declining steeply.

Permafrost—soil that usually remains frozen year-round—has been thawing. Buildings constructed atop permafrost are collapsing, trees are toppling and roads are buckling. In addition to disrupting daily life for local residents, thawing soils also can release large quantities of heat-trapping gases into the atmosphere. When the organic matter that has been locked in permafrost for thousands of years thaws, bacteria break it down into carbon dioxide (if oxygen is present) or methane (if it is not). Arctic permafrost contains about twice as much carbon as the atmosphere holds now, so widespread thaw could greatly exacerbate global warming—which would lead to even faster thaw. Today's computer models do not adequately capture the impacts of thawing permafrost, again contributing to substantial underestimates of future global warming.

The third type of previously permanent Arctic ice is the frozen water on land, including glaciers and Greenland's enormous ice cap, which is more than a mile thick. Losing this ice has dire consequences worldwide because, unlike melting sea ice, the runoff into the ocean raises sea level directly. During the summer of 2016 the total mass of Greenland's ice sheet (estimated using satellite measurements of how it affects the earth's gravity) reached the smallest value since satellite observations began in 2002. The levels are also lower than any amount going back to the late 1950s, when the mass was estimated in other ways. A recent investigation suggests that the accelerating ice melt on Greenland's surface is hastened by warming effects linked to dwindling sea ice.

HOTTER, WETTER AIR

REDUCED SEA ICE and rapid Arctic warming have other far-reaching effects. The combination may alter upper-level winds so that they carry additional heat and moisture from southerly latitudes toward the North Pole. In 2012 the then record Greenland surface melt resulted from an unusually strong and persistent ridge of high pressure in the atmosphere—a so-called blocking high. It brought not only heat and moisture from the south but also soot from wildfires around the Northern Hemisphere. That soot darkens ice and snow surfaces (lowering its albedo, or reflectivity), which then absorb more of the sun's energy, accelerating melt—another one of those vicious cycles.

Strong Arctic warming can lead to prolonged heat waves, stalled hurricanes, relentless rains and more intense fire seasons.

Blocking patterns—large eddies in the jet stream—near Greenland appear to be occurring more often in recent decades, especially in summer months, which probably accounts for at least part of the upward trend in melting. Ice-mass loss during the summer of 2016 was the third highest on record, trailing 2010 and 2012. New work by my colleagues and me suggests that the greater number of blocking highs most likely is tied to global warming. Computer simulations struggle to form and break down blocks realistically, however, so it is hard to say how these patterns may act in the future.

Other Arctic behavior is equally outlandish. During the past two winters record-breaking heat waves near the North Pole were surpassed by more record-breaking heat waves. Waning and thinning sea ice is part of the cause, providing less of a barrier for the ocean's heat to enter the air. The jet stream's wild north-south swings also brought record pulses of warmth and moisture to far northern latitudes. Scientists and Arctic residents often fail to appreciate the strong effects that the extra moisture can create. For starters, water vapor is a greenhouse gas, so in a dry Arctic winter atmosphere, a little more moisture can trap substantially more heat. Moreover, when that moisture condenses into clouds, it releases latent heat, further warming the air. Finally, more clouds trap more warmth below them, one more factor contributing to the Arctic meltdown.

STUCK IN EXTREMES

ALTHOUGH WE HAVE MORE TO LEARN, it has become clear that rapid Arctic change is under way, the most dramatic in hu-

© 2018 Scientific American

MELTWATER atop a disintegrating ice cap in Svalbard, Norway, pours into the sea.

GETTY IMAGES

man history. Given this stark reality, atmospheric scientists are trying to pin down the effects that the changing Arctic may have on people and ecosystems around the world so society can decide how to react and prepare for the future.

An obvious example of global effects is coastal flooding. According to a new report by the Union of Concerned Scientists, about 170 U.S. coastal communities will experience chronic inundation within 20 years. By the end of this century, if nations keep emitting carbon dioxide as they have been, most large coastal cities around the world will face regular, disruptive flooding. This eerily prescient report was published just weeks before Hurricanes Harvey, Irma and Maria capped the U.S.'s most destructive and costly hurricane season ever.

There is also mounting evidence that strong warm-ups in the Arctic's lower atmosphere can affect winds in the jet stream and even higher, in the stratosphere, home to the polar vortex wind pattern that circles the Arctic. Northward peaks and southward valleys in the bending jet stream generate the high- and low-pressure centers we see on weather maps as a capital H or L. The bends control our weather in the Northern Hemisphere. But if extremely large bends occur more often, we can expect to see an uptick in extreme conditions where billions of people live. That is because large bends in the jet stream tend to progress more slowly from west to east, causing the weather systems they create to hang around longer. Think prolonged heat waves, relentless rains and stalled tropical storms like Hurricane Harvey, which swamped Houston in August 2017—as well as more intense fire seasons like the one in California, also in 2017.

Large waves in the jet stream, along with strong Arctic warming, can disrupt the polar vortex, prolonging deadly deep freezes or parades of snowstorms, including the long stretch of severe cold that gripped the northern U.S. in early January of this year. A polar vortex collapse can also perpetuate wild jet-stream swings that deliver crazy heat waves to Alaska and the far north, creating yet another vicious cycle that accelerates Arctic warming. Some studies suggest that Arctic warming is closely connected with these wavy patterns; others say proof of the connection is still tenuous. Research on this hot topic is advancing quickly.

A rapidly warming Arctic most likely will significantly alter habitats on land and at sea. Already, as sea ice recedes, plankton blooms have appeared in new areas during new seasons, enticing species of fish from lower latitudes to move into Arctic waters while native fish disappear. Earlier spring snowmelt across high latitudes has led to earlier greening of the tundra and earlier hatching of insects; migrating birds, which take their cues from the length of daylight hours, may arrive at Arctic food sites too late for the feast. Peoples of the Arctic are feeling the impacts, too; melting ice is keeping them from traditional hunting grounds and even forcing them to uproot towns threatened by coastal erosion from high storm waves in areas that used to be protected by ice along the shore. At the same time, big countries and big companies are swooping in to search for newly accessible natural resources, as tensions rise over who can claim which parts of the vast, rich seabed as theirs.

The revelation my colleagues and I had at the Big Sky retreat replays in my head every time another long-lived weather pattern wreaks havoc or another Arctic record is broken. Now my neighbors are catching on. Polls suggest that most Americans think that the loss of Arctic ice and the jet stream—which has fast become a household term—are conspiring to create weird weather. The old Arctic may have been ruthless, but it was stable. The new Arctic is less predictable and may be changing in ways that are irreversible, with ripple effects on life around the globe.

Are these impacts still avoidable? Yes and no. Because the climate's response lags behind the increases in greenhouse gas concentrations and because carbon dioxide has a very long lifetime in the atmosphere, future change is already baked into the system. But the magnitude and pace can be reduced if society moves quickly to slow emissions and if methods can be developed to extract carbon from the atmosphere in large quantities. Progress on both these fronts is rapid, though likely too little, too late, to preserve the earth and the Arctic as we have known them. Prepare for the unexpected. SA

Jennifer A. Francis has been a research professor in the department of marine and coastal sciences at Rutgers University since 1994. She specializes in Arctic climate change and its links to weather worldwide.

MORE TO EXPLORE

Arctic Matters: The Global Connection to Changes in the Arctic. National Research Council. National Academies Press, 2015.

Amplified Arctic Warming and Mid-latitude Weather: New Perspectives on Emerging Connections. Jennifer A. Francis, Stephen J. Vavrus and Judah Cohen in *WIREs Climate Change,* Vol. 8, No. 5, Article. No. e474; September/October 2017.

FROM OUR ARCHIVES

The Jet Stream Is Getting Weird. Jeff Masters; December 2014.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

TOP SCIENCE STORIES OF 2018
Where Science Meets Society

SCIENCE OF INEQUALITY

A RIGGED ECONOMY

And what we can do about it *By Joseph E. Stiglitz*

AMERICANS ARE USED TO THINKING THAT THEIR NATION is special. In many ways, it is: the U.S. has by far the most Nobel Prize winners, the largest defense expenditures (almost equal to the next 10 or so countries put together) and the most billionaires (twice as many as China, the closest competitor). But some examples of American Exceptionalism should not make us proud. By most accounts, the U.S. has the highest level of economic inequality among developed countries. It has the world's greatest per capita health expenditures yet the lowest life expectancy among comparable countries. It is also one of a few developed countries jostling for the dubious distinction of having the lowest measures of equality of opportunity.

The notion of the American Dream—that, unlike old Europe, we are a land of opportunity—is part of our essence. Yet the numbers say otherwise. The life prospects of a young American depend more on the income and education of his or her parents than in almost any other advanced country. When poor-boy-makes-good anecdotes get passed around in the media, that is precisely because such stories are so rare.

Things appear to be getting worse, partly as a result of forces, such as technology and globalization, that seem beyond our control, but most disturbingly because of those within our command. It is not the laws of nature that have led to this dire situation: it is the laws of humankind. Markets do not exist in a vacuum: they are shaped by rules and regulations, which can be designed to favor one group over another. President Donald Trump was right in saying that the system is rigged—by those in the inherited plutocracy of which he himself is a member. And he is making it much, much worse.

America has long outdone others in its level of inequality, but in the past 40 years it has reached new heights. Whereas the income share of the top 0.1 percent has more than quadrupled and that of the top 1 percent has almost doubled, that of the bottom 90 percent has declined. Wages at the bottom, adjusted for inflation, are about the same as they were some 60 years ago! In fact, for those with a high school education or less, incomes have fallen over recent decades. Males have been particularly hard hit, as the U.S. has moved away from manufacturing industries into an economy based on services.

DEATHS OF DESPAIR

WEALTH IS even less equally distributed, with just three Americans having as much as the bottom 50 percent—testimony to how much money there is at the top and how little there is at the bottom. Families in the bottom 50 percent hardly have the cash reserves to meet an emergency. Newspapers are replete with stories of those for whom the breakdown of a car or an illness starts a downward spiral from which they never recover.

In significant part because of high inequality, U.S. life expectancy,

IN BRIEF

Economic inequality is higher in the U.S. than in virtually all other advanced countries.

The American political system, coupled with high initial inequality, gave the moneyed enough political influence to change laws to benefit themselves, further exacerbating inequality.

Breaking this feedback loop by curbing the power of money in politics is essential to reducing inequality and restoring hope.

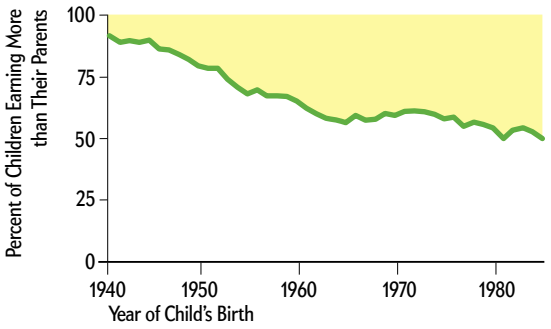
Illustration by Andrea Ucini

© 2018 Scientific American

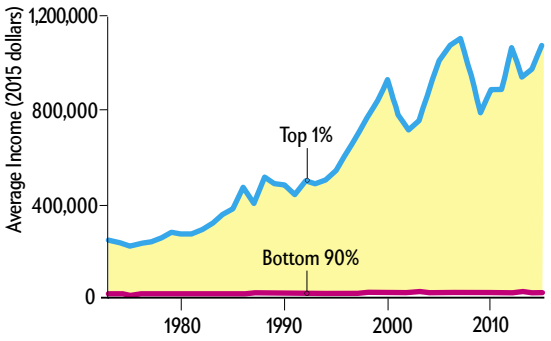
FADING OF THE AMERICAN DREAM

Contrary to popular belief, equality of opportunity in the U.S. is lower than in most advanced countries—and it is declining. A 2017 report by economist Raj Chetty and others indicates that an American born in 1940 was almost certain to become more prosperous than his or her parents. Someone born in 1980 is just as likely to be worse off, however. Declining equality of opportunity stems in large part from the high cost of higher education, coupled with spiraling economic inequality. Statistics from the World Inequality Database show that since about 1970 the income of the top 1 percent, corrected for inflation, has quadrupled, whereas that of the bottom 90 percent has stagnated. Men with only high school degrees have seen their incomes drop.

The American Dream Is Fading for Many …



… as the Income Gap Widens in the U.S.



exceptionally low to begin with, is experiencing sustained declines. This in spite of the marvels of medical science, many advances of which occur right here in America and which are made readily available to the rich. Economist Ann Case and 2015 Nobel laureate in economics Angus Deaton describe one of the main causes of rising morbidity—the increase in alcoholism, drug overdoses and suicides—as "deaths of despair" by those who have given up hope.

Defenders of America's inequality have a pat explanation. They refer to the workings of a competitive market, where the laws of supply and demand determine wages, prices and even interest rates—a mechanical system, much like that describing the physical universe. Those with scarce assets or skills are amply rewarded, they argue, because of the larger contributions they make to the economy. What they get merely represents what they have contributed. Often they take out less than they contributed, so what is left over for the rest is that much more.

This fictional narrative may at one time have assuaged the guilt of those at the top and persuaded everyone else to accept this sorry state of affairs. Perhaps the defining moment exposing the lie was the 2008 financial crisis, when the bankers who brought the global economy to the brink of ruin with predatory lending, market manipulation and various other antisocial practices walked away with millions of dollars in bonuses just as millions of Americans lost their jobs and homes and tens of millions more worldwide suffered on their account. Virtually none of these bankers were ever held to account for their misdeeds.

I became aware of the fantastical nature of this narrative as a schoolboy, when I thought of the wealth of the plantation owners, built on the backs of slaves. At the time of the Civil War, the market value of the slaves in the South was approximately half of the region's total wealth, including the value of the land and the physical capital—the factories and equipment. The wealth of at least this part of this nation was not based on industry, innovation and commerce but rather on exploitation. Today we have replaced this open exploitation with more insidious forms, which have intensified since the Reagan-Thatcher revolution of the 1980s. This exploitation, I will argue, is largely to blame for the escalating inequality in the U.S.

After the New Deal of the 1930s, American inequality went into decline. By the 1950s inequality had receded to such an extent that another Nobel laureate in economics, Simon Kuznets, formulated what came to be called Kuznets's law. In the early stages of development, as some parts of a country seize new opportunities, inequalities grow, he postulated; in the later stages, they shrink. The theory long fit the data—but then, around the early 1980s, the trend abruptly reversed.

EXPLAINING INEQUALITY

ECONOMISTS HAVE put forward a range of explanations for why inequality has in fact been increasing in many developed countries. Some argue that advances in technology have spurred the demand for skilled labor relative to unskilled labor, thereby depressing the wages of the latter. Yet that alone cannot explain why even skilled labor has done so poorly over the past two decades, why average wages have done so badly and why matters are so much worse in the U.S. than in other developed nations. Changes in technology are global and should affect all advanced economies in the same way. Other economists blame globalization itself, which has weakened the power of workers. Firms can and do move abroad unless demands for higher wages are curtailed. But again, globalization has been integral to all advanced economies. Why is its impact so much worse in the U.S.?

The shift from a manufacturing to a service-based economy is partly to blame. At its extreme—a firm of one person—the service economy is a winner-takes-all system. A movie star makes millions, for example, whereas most actors make a pittance. Overall, wages are likely to be far more widely dispersed in a service economy than in one based on manufacturing, so the transition contributes to greater inequality. This fact does not explain, however, why the average wage has not improved for decades. Moreover, the shift to the service sector is happening in most other advanced countries: Why are matters so much worse in the U.S.?

Again, because services are often provided locally, firms have more market power: the ability to raise prices above what would prevail in a competitive market. A small town in rural America may have only one

SOURCES: "THE FADING AMERICAN DREAM: TRENDS IN ABSOLUTE INCOME MOBILITY SINCE 1940," BY RAJ CHETTY ET AL., IN *SCIENCE*, VOL. 356; APRIL 28, 2017 (*child-parent wealth comparison*); WORLD INEQUALITY DATABASE (*90% versus 1% wealth trend data*)

Graphics by Jen Christiansen

© 2018 Scientific American

authorized Toyota repair shop, which virtually every Toyota owner is forced to patronize. The providers of these local services can raise prices over costs, increasing their profits and the share of income going to owners and managers. This, too, increases inequality. But again, why is U.S. inequality practically unique?

In his celebrated 2013 treatise *Capital in the Twenty-First Century,* French economist Thomas Piketty shifts the gaze to capitalists. He suggests that the few who own much of a country's capital save so much that, given the stable and high return to capital (relative to the growth rate of the economy), their share of the national income has been increasing. His theory has, however, been questioned on many grounds. For instance, the savings rate of even the rich in the U.S. is so low, compared with the rich in other countries, that the increase in inequality should be lower here, not greater.

An alternative theory is far more consonant with the facts. Since the mid-1970s the rules of the economic game have been rewritten, both globally and nationally, in ways that advantage the rich and disadvantage the rest. And they have been rewritten further in this perverse direction in the U.S. than in other developed countries—even though the rules in the U.S. were already less favorable to workers. From this perspective, increasing inequality is a matter of choice: a consequence of our policies, laws and regulations.

In the U.S., the market power of large corporations, which was greater than in most other advanced countries to begin with, has increased even more than elsewhere. On the other hand, the market power of workers, which started out less than in most other advanced countries, has fallen further than elsewhere. This is not only because of the shift to a service-sector economy—it is because of the rigged rules of the game, rules set in a political system that is itself rigged through gerrymandering, voter suppression and the influence of money. A vicious spiral has formed: economic inequality translates into political inequality, which leads to rules that favor the wealthy, which in turn reinforces economic inequality.

FEEDBACK LOOP

POLITICAL SCIENTISTS have documented the ways in which money influences politics in certain political systems, converting higher economic inequality into greater political inequality. Political inequality, in its turn, gives rise to more economic inequality as the rich use their political power to shape the rules of the game in ways that favor them—for instance, by softening antitrust laws and weakening unions. Using mathematical models, economists such as myself have shown that this two-way feedback loop between money and regulations leads to at least two stable points. If an economy starts out with lower inequality, the political system generates rules that sustain it, leading to one equilibrium situation. The American system is the other equilibrium—and will continue to be unless there is a democratic political awakening.

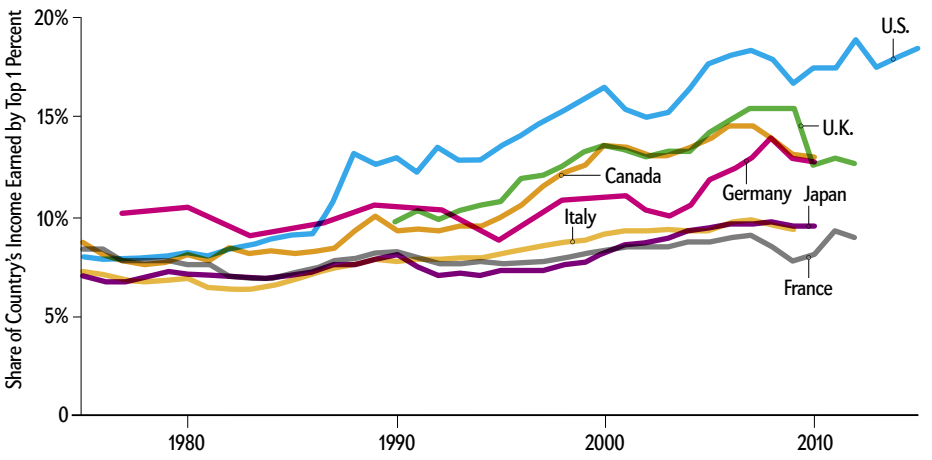
An account of how the rules have been shaped must begin with antitrust laws, first enacted 128 years ago in the U.S. to prevent the agglomeration of market power. Their enforcement has weakened—at a time when, if anything, the laws themselves should have been strengthened. Technological changes have concentrated market power in the hands of a few global players, in part because of so-called network effects: you are far more likely to join a particular social network or use a certain word processor if everyone you know is already using it. Once established, a firm such as Facebook or Microsoft is hard to dislodge. Moreover, fixed costs, such as that of developing a piece of software, have increased as compared with marginal costs—that of duplicating the software. A new entrant has to bear all these fixed costs up front, and if it does enter, the rich incumbent can respond by lowering prices drastically. The cost of making an additional e-book or photo-editing program is essentially zero.

In short, entry is hard and risky, which gives established firms with deep war chests enormous power to crush competitors and ultimately raise prices. Making matters worse, U.S. firms have been innovative not only in the products they make but in thinking of ways to extend and amplify their market power. The European Commission has imposed

Joseph E. Stiglitz is a University Professor at Columbia University and Chief Economist at the Roosevelt Institute. He received the Nobel prize in economics in 2001. Stiglitz chaired the Council of Economic Advisers from 1995–1997, during the Clinton administration, and served as the chief economist and senior vice president of the World Bank from 1997–2000. He chaired the United Nations commission on reforms of the international financial system in 2008–2009. His latest authored book is *Globalization and Its Discontents Revisited* (2017).

GLOBAL INEQUALITY TRENDS

Inequality has increased in most advanced countries because of factors such as globalization, technological change and the shift to a service-based economy. It has grown fastest in the U.S., however, according to the World Inequality Database. That is because rules have been rewritten to make them more favorable for the rich, while being disadvantageous to everyone else. Large companies have been allowed to accrue more power over the market, whereas the influence of workers has shrunk. Taxation and other policies have consistently favored the wealthy.



SOURCES: *ECONOMIC REPORT OF THE PRESIDENT*, JANUARY 2017; WORLD INEQUALITY DATABASE

© 2018 Scientific American

fines of billions of dollars on Microsoft and Google and ordered them to stop their anticompetitive practices (such as Google privileging its own comparison shopping service). In the U.S., we have done too little to control concentrations of market power, so it is not a surprise that it has increased in many sectors.

Rigged rules also explain why the impact of globalization may have been worse in the U.S. A concerted attack on unions has almost halved the fraction of unionized workers in the nation, to about 11 percent. (In Scandinavia, it is roughly 70 percent.) Weaker unions provide workers less protection against the efforts of firms to drive down wages or worsen working conditions. Moreover, U.S. investment treaties such as the North Atlantic Free Trade Agreement—treaties that were sold as a way of preventing foreign countries from discriminating against American firms—also protect investors against a tightening of environmental and health regulations abroad. For instance, they enable corporations to sue nations in private international arbitration panels for passing laws that protect citizens and the environment but threaten the multinational company's bottom line. Firms like these provisions, which enhance the credibility of a company's threat to move abroad if workers do not temper their demands. In short, these investment agreements weaken U.S. workers' bargaining power even further.

LIBERATED FINANCE

MANY OTHER changes to our norms, laws, rules and regulations have contributed to inequality. Weak corporate governance laws have allowed chief executives in the U.S. to compensate themselves 361 times more than the average worker, far more than in other developed countries. Financial liberalization—the stripping away of regulations designed to prevent the financial sector from imposing harms, such as the 2008 economic crisis, on the rest of society—has enabled the finance industry to grow in size and profitability and has increased its opportunities to exploit everyone else. Banks routinely indulge in practices that are legal but should not be, such as imposing usurious interest rates on borrowers or exorbitant fees on merchants for credit and debit cards and creating securities that are designed to fail. They also frequently do things that are illegal, including market manipulation and insider trading. In all of this, the financial sector has moved money away from ordinary Americans to rich bankers and the banks' shareholders. This redistribution of wealth is an important contributor to American inequality.

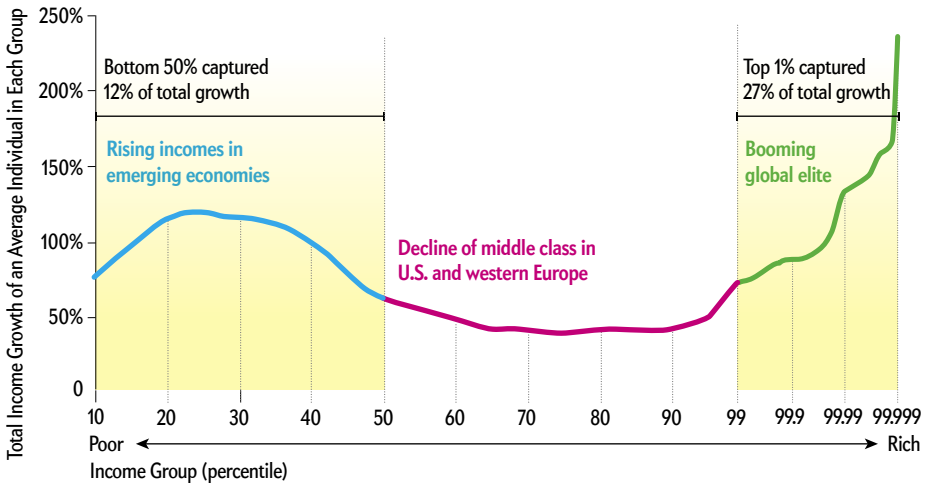
Other means of so-called rent extraction—the withdrawal of income from the national pie that is incommensurate with societal contribution—abound. For example, a legal provision enacted in 2003 prohibited the government from negotiating drug prices for Medicare—a gift of some $50 billion a year or more to the pharmaceutical industry. Special favors, such as extractive industries' obtaining public resources such as oil at below fair-market value or banks' getting funds from the Federal Reserve at near-zero interest rates (which they relend at high interest rates), also amount to rent extraction. Further exacerbating inequality is favorable tax treatment for the rich. In the U.S., those at the top pay a smaller fraction of their income in taxes than those who are much poorer—a form of largesse that the Trump administration has just worsened with the 2017 tax bill.

Some economists have argued that we can lessen inequality only by giving up on growth and efficiency. But recent research, such as work done by Jonathan Ostry and others at the International Monetary Fund, suggests that economies with greater equality perform better, with higher growth, better average standards of living and greater stability. Inequality in the extremes observed in the U.S. and in the manner generated there actually damages the economy. The exploitation of market power and the variety of other distortions I have described, for instance, makes markets less efficient, leading to underproduction of valuable goods such as basic research and overproduction of others, such as exploitative financial products.

Moreover, because the rich typically spend a smaller fraction of

UNEVEN DISTRIBUTION OF GLOBAL GROWTH

Globalization has benefited millions of the poor in emerging economies, particularly in China. Data compiled by economist Branko Milanovic and displayed in the *World Inequality Report 2018* demonstrate, however, that between 1980 and 2016, the steepest gains went to the world's top 1 percent, which captured more than a quarter of the growth in the global economy. In early 2018 Oxfam International reported that just 42 individuals have as much wealth as the bottom 50 percent put together. The middle classes in the U.S. and western Europe have benefited the least from global growth, as have the world's poorest.



SOURCES: *WORLD INEQUALITY REPORT 2018*. WORLD INEQUALITY LAB, 2017; BRANKO MILANOVIC

© 2018 Scientific American

their income on consumption than the poor, total or "aggregate" demand in countries with higher inequality is weaker. Societies could make up for this gap by increasing government spending—on infrastructure, education and health, for instance, all of which are investments necessary for long-term growth. But the politics of unequal societies typically puts the burden on monetary policy: interest rates are lowered to stimulate spending. Artificially low interest rates, especially if coupled with inadequate financial market regulation, can give rise to bubbles, which is what happened with the 2008 housing crisis.

It is no surprise that, on average, people living in unequal societies have less equality of opportunity: those at the bottom never get the education that would enable them to live up to their potential. This fact, in turn, exacerbates inequality while wasting the country's most valuable resource: Americans themselves.

RESTORING JUSTICE

MORALE IS LOWER in unequal societies, especially when inequality is seen as unjust, and the feeling of being used or cheated leads to lower productivity. When those who run gambling casinos or bankers suffering from moral turpitude make a zillion times more than the scientists and inventors who brought us lasers, transistors and an understanding of DNA, it is clear that something is wrong. Then again, the children of the rich come to think of themselves as a class apart, entitled to their good fortune, and accordingly more likely to break the rules necessary for making society function. All of this contributes to a breakdown of trust, with its attendant impact on social cohesion and economic performance.

There is no magic bullet to remedy a problem as deep-rooted as America's inequality. Its origins are largely political, so it is hard to imagine meaningful change without a concerted effort to take money out of politics—through, for instance, campaign finance reform. Blocking the revolving doors by which regulators and other government officials come from and return to the same industries they regulate and work with is also essential.

Beyond that, we need more progressive taxation and high-quality federally funded public education, including affordable access to universities for all, no ruinous loans required. We need modern competition laws to deal with the problems posed by 21st-century market power and stronger enforcement of the laws we do have. We need labor laws that protect workers and their rights to unionize. We need corporate governance laws that curb exorbitant salaries bestowed on chief executives, and we need stronger financial regulations that will prevent banks from engaging in the exploitative practices that have become their hallmark. We need better enforcement of antidiscrimination laws: it is unconscionable that women and minorities get paid a mere fraction of what their white male counterparts receive. We also need more sensible inheritance laws that will reduce the intergenerational transmission of advantage and disadvantage.

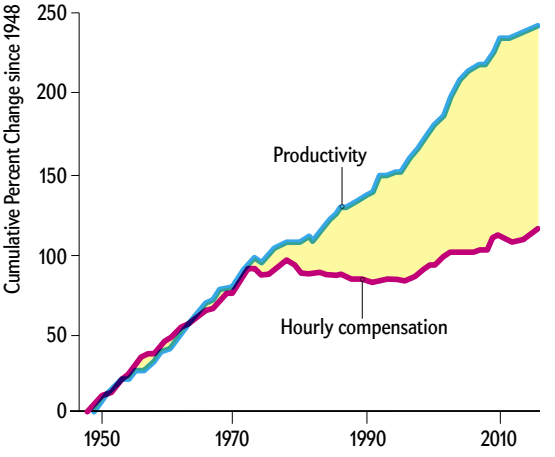
The basic perquisites of a middle-class life, including a secure old age, are no longer attainable for most Americans. We need to guarantee access to health care. We need to strengthen and reform retirement programs, which have put an increasing burden of risk management on workers (who are expected to manage their portfolios to guard simultaneously against the risks of inflation and market collapse) and opened them up to exploitation by our financial sector (which sells them products designed to maximize bank fees rather than retirement security). Our mortgage system was our Achilles' heel, and we have not really fixed it. With such a large fraction of Americans living in cities, we have to have urban housing policies that ensure affordable housing for all.

It is a long agenda—but a doable one. When skeptics say it is nice but not affordable, I reply: We cannot afford to *not* do these things. We are already paying a high price for inequality, but it is just a down payment on what we will have to pay if we do not do something—and quickly. It is not just our economy that is at stake; we are risking our democracy.

As more of our citizens come to understand why the fruits of economic progress have been so unequally shared, there is a real danger that they will become open to a demagogue blaming the country's problems on others and making false promises of rectifying "a rigged system." We are already experiencing a foretaste of what might happen. It could get much worse. SA

WIDENING WAGE GAP

Since about 1980 the productivity of American workers has doubled, according to Josh Bivens and others at the Economic Policy Institute. But wages for production and nonsupervisory workers have stagnated, with virtually all the gains from increased productivity going to investors and owners. Salaries for the top 1 percent, including corporate executives and finance professionals, have, however, gone up—by more than 150 percent between 1979 and 2012. The increasing wage gap plays a significant role in spiraling inequality.



SOURCES: *RAISING AMERICA'S PAY: WHY IT'S OUR CENTRAL ECONOMIC POLICY CHALLENGE*, BY JOSH BIVENS ET AL. ECONOMIC POLICY INSTITUTE, JUNE 4, 2014; *THE STATE OF WORKING AMERICA*, BY LAWRENCE MISHEL, JOSH BIVENS, ELISE GOULD AND HEIDI SHIERHOLZ, 12TH EDITION. ILR PRESS, 2012

MORE TO EXPLORE

The Price of Inequality: How Today's Divided Society Endangers Our Future. Joseph E. Stiglitz. W. W. Norton, 2012.

The Great Divide: Unequal Societies and What We Can Do about Them. Joseph E. Stiglitz. W. W. Norton, 2015.

Rewriting the Rules of the American Economy: An Agenda for Growth and Shared Prosperity. Joseph E. Stiglitz. W. W. Norton, 2015.

Globalization and Its Discontents Revisited: Anti-globalization in the Era of Trump. Joseph E. Stiglitz. W. W. Norton, 2017.

FROM OUR ARCHIVES

The Threat of Inequality. Angus Deaton; September 2016.

scientificamerican.com/magazine/sa

© 2018 Scientific American

THE FUTURE OF MONEY

THE WORLD BITCOIN CREATED

THE FIRST BIG DIGITAL CURRENCY GAVE US A GLIMPSE OF A NEW ECONOMIC ORDER—ONE THAT RAISES MORE QUESTIONS THAN ANSWERS

BY JOHN PAVLUS

BITCOIN. CRYPTOCURRENCIES. SMART CONTRACTS. MANY PEOPLE have now heard of the rapidly changing ecosystem of financial technology, but few have wrapped their heads around it. Hundreds of central banks and corporations are incubating a game-changing technology called blockchain—and investors are betting billions on it. Yet only 24 percent of global financial services professionals surveyed in 2017 by PricewaterhouseCoopers (PwC) described themselves as "extremely" or "very" familiar with it. Much of the public is unsure if any of this is legal, if they understand it at all. Evangelists say it has the power to upend entire economic systems; others, such as Emin Gün Sirer, a blockchain researcher at Cornell University, warn that while the technical core is "fascinating and disruptive, there's also a lot of hokum out there." How to parse the nuance—or get a handle on what a blockchain is?

It all starts with Satoshi Nakamoto, the world's most reclusive pseudonymous billionaire. In October 2008 Nakamoto published a paper via an obscure Internet mailing list detailing a design for the world's first blockchain: a public database distributed and synchronized every 10 minutes across thousands of computers, accessible to anyone and yet hackable by no one. Its purpose? To provide a decentralized, bulletproof record of exchange for a new digital currency Nakamoto called Bitcoin.

Until that point, the trouble with "peer-to-peer electronic cash" was that nobody could reliably prevent you from spending it twice. Blockchain technology changed all that by inscribing every transfer of Bitcoin into a "distributed ledger"—a kind of digital spreadsheet that, thanks to the laws of mathematics and cryptography, was more inviolable than carving it in stone. The *Economist* dubbed it "the trust machine."

The technology that underpins Bitcoin quickly outgrew it, driving a frenetic period of innovation. Think of blockchain as a scaffolding that can hold any data that need secure provenance: financial histories, ownership documents, proofs of identity. This "worldwide ledger"—as Don Tapscott, co-author of *Blockchain Revolution,* calls it—is a blank slate. But the technology, imperfect as it is, can be tapped for evil, too, and some are pumping the brakes on the frenzy. Here's a guide to the digital landscape that Satoshi Nakamoto—whoever he is—has thrust before us.

John Pavlus is a writer and filmmaker focusing on science, technology and design. His work has appeared in *Bloomberg Businessweek, MIT Technology Review,* and *The Best American Science and Nature Writing* series. He lives in Portland, Ore.

CORE CONCEPTS

CRYPTOCURRENCY A form of digital currency that relies on the mathematics of cryptography to control how and when units of the currency are created and to ensure secure transfer of funds.

PEER-TO-PEER (P2P) NETWORK A web of computers linked in a decentralized way, such that any computer can communicate directly with any other without going through a central server or other administrator. Napster, the network for sharing music files that launched in the late 1990s, popularized the concept.

NODE A computer connected to a P2P network. The Bitcoin network currently has thousands of nodes spread across the globe.

DISTRIBUTED LEDGER A list of recorded, time-stamped transactions that is simultaneously broadcast, copied and verified via consensus across many different computers in a P2P network. If every node in the network has an identical copy of the ledger, falsified entries or corrupted versions can be easily detected.

BLOCK A grouping of individual transaction records on a blockchain. On the Bitcoin network, new blocks are added to the chain every 10 minutes.

HASHING A cryptographic method that uses a mathematical function to condense any amount of data into a unique string of alphanumeric characters of a certain fixed length—called a hash value. This creates an easily verifiable digital fingerprint for the hashed data. If even a single bit of the original data is changed or corrupted, the fingerprint that emerges from the hash function will be drastically different, making it easy to detect errors or tampering. Hashes are also "one-way"—the data cannot be reassembled or extracted from the fingerprint.

MINING The process by which nodes of a cryptocurrency network compete to securely add new blocks of transactions to a blockchain. Units of the currency are the reward—and hence, a financial incentive to ensure security. Mining involves downloading the latest version of the blockchain's transactions for verification, then using brute-force computation to randomly search for the solution to a difficult mathematical puzzle created via hashing. The first node to discover the correct solution "mines" that block, adding it to the blockchain and claiming the reward associated with it. Humans control nodes, but the competition has nothing to do with skill: simply, the more raw computing power a miner applies toward the solution, the more likely he or she is to find it—a process called proof of work.

Illustrations by Peter and Maria Hoey

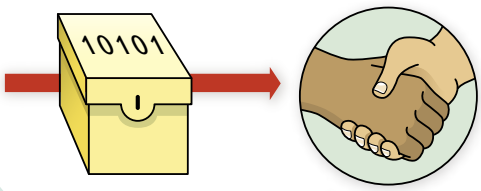
© 2018 Scientific American

HOW BLOCKCHAIN WORKS

How does digital currency—or any data—reliably pass back and forth on a decentralized network full of strangers that don't have a reason to trust one another at all? By generating a permanent ledger of transactions that can't be changed by any single network member.

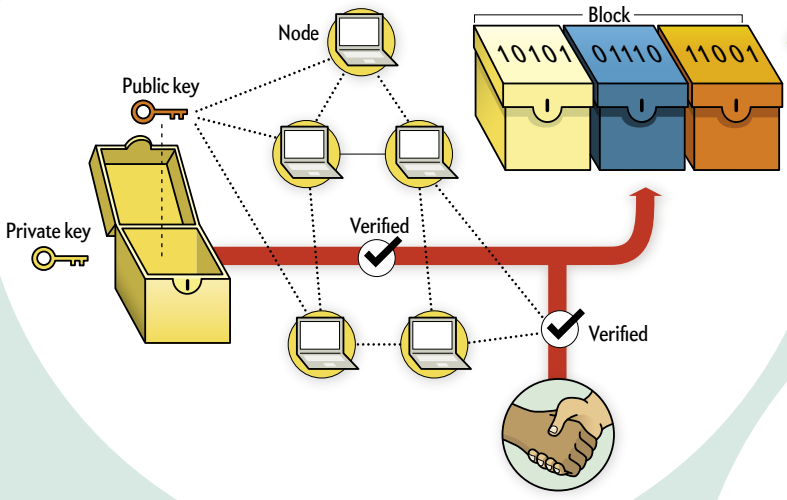
1

A blockchain transaction begins with one party agreeing to send data to another. These data could be anything. But because the point of a blockchain is to create a permanent, verifiable record of exchange, the data usually represent some valuable asset. Common examples: units of a cryptocurrency or other financial instrument; contracts, deeds or records of ownership; medical information or other identity data.



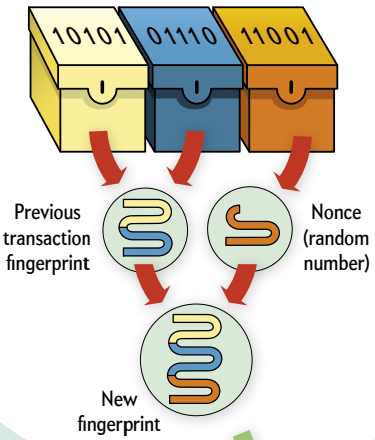
2

The transaction is broadcast for verification to a peer-to-peer network of computers operating the blockchain. Every node on the network is equipped with a procedure for verifying whether the transaction is valid or not. (In a Bitcoin transaction, for example, the network would verify whether those paying actually have the amount of Bitcoins they say they do.) Once the network has reached a consensus, algorithms package up the validated transaction with other recent transactions into a block.



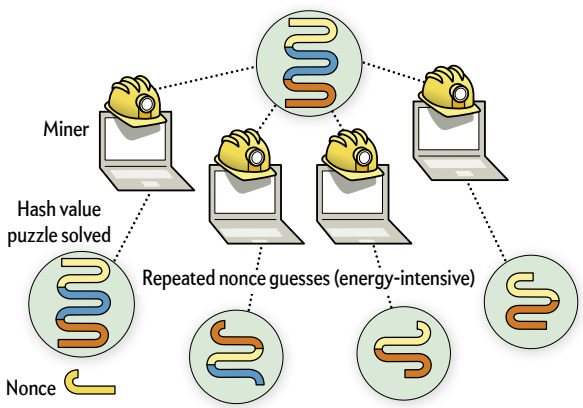
3

Software creates a "fingerprint" for the new block by hashing the data inside it, together with two other pieces of information: the fingerprint of the preceding block and a random number called a nonce.



4

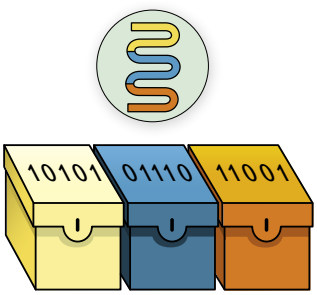
Special nodes called miners begin competing with one another for the right to add the new block to the blockchain. Their computers perform a tedious set of hash-based calculations over and over again by trial and error, hoping to generate a solution that satisfies an arbitrary rule defined by the network. (On the Bitcoin blockchain, the miners are searching for solutions—or "hash values"—that have a particular number of zeros at the beginning.) Whoever is first to complete this proof-of-work process and find the matching solution successfully "mines" that block, earning a financial reward.



AS AN ALTERNATIVE: Proof-of-work mining is energy-intensive, so some new blockchains are doing away with it, instead using a preapproved network of "validator" nodes who can notarize transactions via an alternative process called proof of stake. Because this process doesn't rely on difficult hashing calculations, it uses much less computing power (and much less electricity).

5

The validated block is added to the blockchain with a digital fingerprint that also mathematically encodes the validated fingerprints of every block preceding it. These nested fingerprints make the blockchain increasingly secure with every new block that gets added because altering a single bit of information anywhere in the blockchain would drastically change not only the fingerprint of that particular block but every subsequent one in the chain as well.



© 2018 Scientific American

BLOCKCHAIN DEMYSTIFIED: FREQUENTLY ASKED QUESTIONS ABOUT A RAPIDLY EXPANDING TOPIC

1. ARE BITCOIN AND BLOCKCHAIN THE SAME?

No, but it's easy to get them confused because they both came into public awareness in 2008, when Satoshi Nakamoto published his paper describing how to implement them simultaneously. Bitcoin is one type of cryptocurrency. What people call "blockchain" is a technology that makes Bitcoin possible—an infrastructure that can be used for tracking many types of transactions. Blockchain technology exists without Bitcoin—but not the reverse. Think of Bitcoin as a kind of application that runs "on" the blockchain, much like Web sites run on the Internet.

2. WHERE DOES THE VALUE OF A CRYPTOCURRENCY COME FROM?

Some experts say that a cryptocurrency like Bitcoin has value because of its security (the Bitcoin blockchain has never been hacked—yet) or its mathematically imposed "scarcity" (a fixed supply of 21 million Bitcoins means they can never be devalued by "printing more money"). Others say that they have intrinsic value because mining them is tedious work that makes the network stronger—in other words, there's value in effort. But what about cryptocurrencies that aren't mined? According to Christian Catalini of the Massachusetts Institute of Technology, "value comes from consensus. We all *agree* it has value." In this sense, cryptocurrencies may have more in common with social networks than with central banks. "Money is a way for society to keep track of checks and balances," Catalini says. "If cryptocurrencies end up being a better way to track information," their value is secured—whether they represent a physical asset or just a number.

3. IS THE BLOCKCHAIN A NEW KIND OF INTERNET?

Not quite, because the blockchain itself requires the Internet to support and maintain its peer-to-peer network. It's also important to note that when people talk informally about "the" blockchain, they're almost always referring to the specific system that Nakamoto implemented to support Bitcoin. The Bitcoin blockchain was the first distributed ledger system that didn't require a centralized server or organization to support it. It's still one of the biggest: as of November 2017 it contains more than 130 gigabytes (140 billion bytes) of information, and every new transaction increases its size. But that's still many orders of magnitude smaller than the amount of data on the Internet, which is estimated to be on the yottabyte scale (10^{24}, or septillions of bytes).

4. ARE BLOCKCHAINS EVEN LEGAL?

Yes. But their decentralized nature and association with Bitcoin—which has been used in illegal transactions such as drug and arms sales—can give blockchains an "outlaw" reputation it doesn't necessarily deserve. Blockchains can be used for many different purposes, good or ill, just like Facebook, e-mail or any other Internet technology.

5. HOW ARE CRYPTOCURRENCIES SECURE AND TRUSTWORTHY?

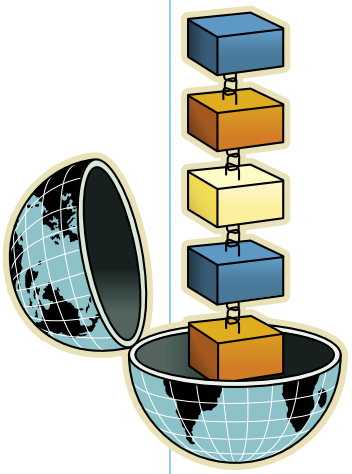
Because they're ultimately nothing but software, the trustworthiness of a cryptocurrency "comes from the code base," says M.I.T. researcher Catalini. Anyone can gin up a cryptocoin and raise funds by selling it through an initial coin offering—even Paris Hilton did it, lending her name to promote an obscure token. But it's no coincidence that the two most popular cryptocurrencies, Bitcoin and Ether, were engineered by computer-programming savants. That said, even coins with impressive technical bona fides can be risky. The DAO—a "decentralized autonomous organization" running on Ethereum that raised over $100 million in 2016—"had a bug" (in Catalini's understated terms) that allowed hackers to make off with $50 million worth of Ether.

© 2018 Scientific American

© 2018 Scientific American

10%

Predicted amount of the world's gross domestic product that will be stored in blockchain-based technology by 2025, according to a 2015 survey report from the World Economic Forum.



WHO IS USING BLOCKCHAIN TECHNOLOGY?

It's not just for cyberlibertarians, and it goes way beyond finance. Here's an incomplete lineup:

- **FINANCIAL INSTITUTIONS:** Global banks and investment institutions are researching and pursuing blockchain projects, sometimes joining forces in consortiums. Since 2012 Ripple has been a thriving, blockchain-based system for settling international transactions among banks. Start-ups such as Bloom intend to deploy blockchains to credit reporting, hoping to end data breaches like the Equifax hack.
- **GOVERNMENTS:** Delaware and Illinois use distributed ledgers for birth certificates. A Vermont law allows blockchain technology to verify the authenticity of legal documents. Dubai integrated blockchains into many of its administrative services, such as obtaining licenses. In 2016 Tunisia began issuing a blockchain-backed version of its digital national currency called the eDinar.
- **TECH ENTREPRENEURS:** The Ethereum network—which was designed to support new applications, rather than just a digital cash ecosystem like Bitcoin—is like an App Store for blockchain start-ups. Hundreds of projects and businesses are running on it. One notable: WePower wants to let households buy and sell renewable energy (from, say, roof-mounted solar panels) directly to one another.
- **COPYRIGHT AND IP HOLDERS:** U.K. musician Imogen Heap started Mycelia, a tech incubator that tracks metadata associated with creative works, cutting out intermediaries like iTunes.
- **NONPROFITS AND AID GROUPS:** The BitGive Foundation is boosting the accountability of philanthropic giving. And the United Nations World Food Program is streamlining how it tracks and delivers assistance to Syrian refugees in Jordan.
- **ACADEMIC INSTITUTIONS:** Forget sheepskins. The Blockcerts project wants to make all manner of academic and professional credentials more trustworthy and shareable.
- **ASSET MANAGERS:** London-based Everledger is targeting the diamond industry by recording the attributes and provenance of each precious stone. Fine wine and art are tracked, too.
- **JOURNALISTS:** To push back against fake news, Civil gives news makers a platform to create ad-free, inalterable journalism that's immune to outside interests (Russia; Facebook) and supported by readers.
- **REGULAR PEOPLE:** For migrant workers who send money to their families back home, using Bitcoin costs less than using Western Union, which is why an estimated 20 percent of international remittances between South Korea and the Philippines now rely on it.

77%

of the global financial services industry is expected to adopt blockchain as part of a production system or process by 2020, according to PwC.

SO, BITCOIN: THE FUTURE OR A FLASH IN THE PAN?

Bitcoin is the world's most popular digital currency. But it's also wildly speculative, and many financial experts point to its legendary volatility: the currency's value has risen more than 10-fold since 2016, but it lost 40 percent of its value in a span of two weeks in September 2017—only to regain (and surpass) it just as quickly. (Who knows what it will be by the time you read this.) To others, the network's technical limitations—it is sluggish at handling transactions—combined with its unsustainable mining costs make it the equivalent of a financial time bomb. "We don't bet on Bitcoin," says Charlie Morris, chief investment officer of NextBlock Global, a firm that invests in blockchain technology.

Bitcoin legitimized the basic economics of a global cryptocurrency. But the next-largest "altcoin" may have more staying power: Ether is less a cashlike currency than a "blockchain asset," as Morris calls it, used to power and secure the Ethereum network. Much like renting virtual servers in Google's "cloud," developers who want to create applications using Ethereum's blockchain must pay for access in tokens of Ether. The more useful Ethereum becomes as a mainstream platform, the more stable and valuable Ether becomes, too. New currencies and platforms are very likely to emerge—the race for prominence has only just begun.

WHY WOULD YOU USE A CRYPTOCURRENCY INSTEAD OF A NATIONAL CURRENCY?

Imagine holding a $100 bill that buys only $50 worth of goods. In Venezuela, where the official currency is crashing in value, that scenario is a reality. "You're losing something like half of the value of your net income every year to hyperinflation," says venture capitalist Morris. "People are thinking: 'How can I stop that?' And they're buying Bitcoin."

Why would a hard-to-understand cryptocurrency with no government guaranteeing its value as "legal tender" seem like a better bet than a more traditional value-holding commodity such as gold? For one thing, converting Venezuelan bolivars into Bitcoin is simply a lot easier for ordinary folks—anyone with access to the Internet can do it. Because Bitcoin has no physical form, you don't have to stash it somewhere unsafe—like a mattress or, in Venezuela's case, a bank. Of course, Bitcoin doesn't have a stable value, either. But while the bolivar has nose-dived, the value of a Bitcoin is at least trending ever upward. In a country where inflation is expected to exceed 2,300 percent in 2018 (according to the International Monetary Fund), it seems like a reasonable risk to take.

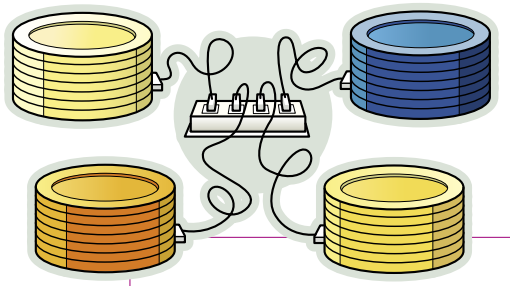
Zimbabweans have the opposite problem. After ditching its own currency for the U.S. dollar, the country now relies on currency imports to run its economy—and it's facing a shortage. Bitcoin is now common enough that it's even accepted by car dealers.

ARE WE FACING THE END OF CASH?

It may seem that printed money is headed for the same fate as newspapers. But experts say that cash is far from dead. "We're still using great piles of paper to pay for things like international shipping of sea containers," says Vinay Gupta, CEO of Mattereum, a legal services firm for smart contracts. "The system is not so broken that people are willing to tear it up." The trouble with Bitcoin and Ether is that while they can function as a store of value or unit of exchange, they're not accepted as legal tender in enough places to compete with cash.

In places such as Kenya, where few people have traditional bank accounts and "mobile money" services such as M-Pesa have made saving and sending money by phone much easier than exchanging physical cash, cryptocurrencies might seem like a natural fit. But mining still requires a lot of processing power—not a common resource in Africa, where inexpensive feature phones outsell smartphones and not many own PCs. The computations required to secure blockchain transactions could, in theory, happen on "your old Nokia SIM card," Gupta says. Still, cold, hard paper won't soon disappear.

3 4 5



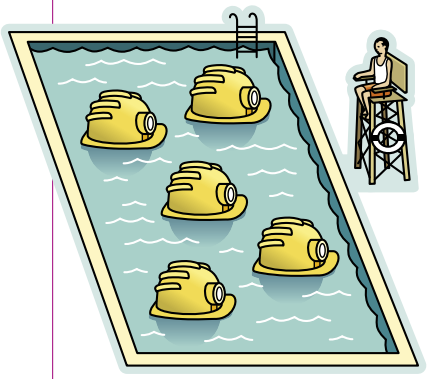
IS THE BLOCKCHAIN A NEW KIND OF INTERNET?

IF CRYPTOCURRENCIES ARE DIGITAL, WHAT POWERS THEM?

Just because cryptocurrencies have no physical attributes doesn't mean that there's no cost to using them. The intentionally effortful process by which new Bitcoins are "mined"—how new transactions are added to the ledger—requires that the entire P2P network cycle through a mind-boggling number of random computations to validate blockchain transactions. All of that processing requires energy.

How much energy? Start with the amount of computation. In late 2017 the Bitcoin network's "hash rate" was around 10 exahashes—that's 10 million trillion calculations—per second. Deriving a precise energy estimate from that figure is impossible because the network, being decentralized, can't account for individual nodes. But credible estimates peg the Bitcoin network's annual electricity consumption at around 27 terawatt-hours—roughly equivalent to that of Ireland. To put that in perspective, producing a year's worth of Bitcoin alone requires the equivalent of burning about 11 million tons of coal, which pours nearly 29 million tons of carbon dioxide into the atmosphere. Fueling Bitcoin by solar power would require harnessing more than half of the entire U.S.'s annual utility-scale solar capacity.

Ethereum's creator, Vitalik Buterin, is currently transitioning the network's blockchain to a different validation mechanism called proof of stake, which doesn't rely on mining at all. Bitcoin's larger, more decentralized network is unlikely to make a similar move anytime soon. But Vinay Gupta, who designed Dubai's blockchain strategy, believes that the same greed that motivates miners to turn kilowatts into cryptocurrency will ultimately spur them to innovate their way out of this scalability problem. Venture capitalist Charlie Morris thinks that as proof-of-stake cryptocurrencies prove their mettle in the market, "mining will become like a little blip in history," he says. "People will say, 'Remember when we all did that—wasn't that ridiculous?' "



WHERE DOES MINING ACTUALLY OCCUR?

71%

of Bitcoin is mined in China; the next most active country is India, at 4 percent. Tip: Don't try mining at home—alone. The task is now dominated by giant mining pools akin to the ones in China, so the chances of a solo node mining a block today is about one in eight million. Lone operators would spend far more on energy bills than they'd get in profits. Want to become a mining hobbyist? Join a public mining pool.

ARE BLOCKCHAINS EVEN LEGAL?

HOW ARE CRYPTOCURRENCIES SECURE AND TRUSTWORTHY?

WHAT DOES THE PUBLIC THINK ABOUT BLOCKCHAIN?

62%

of Americans believe cryptocurrencies are used for illegal purchases or don't know what they're used for at all, according to a 2017 YouGov survey.

59%

of global consumers polled in a 2017 HSBC survey said they had never heard of blockchain technology; 80 percent of those who had heard of the technology still don't understand what it is.

39%

of senior executives at large U.S. companies indicated they had little or no knowledge about blockchain technology, according to a 2017 Deloitte survey.

HOW WILL THIS TECHNOLOGY BE USED IN THE FUTURE?

Anyone building on blockchain technology is, by definition, a futurist. Once distributed ledger technology gets out of its training-wheels phase, what might we create with it?

- **SELF-DRIVING, SELF-OWNING CARS:** Instead of driving for Uber, your car would drive itself while you work or sleep. Blockchain-backed smart contracts could remove middlemen like Uber and Lyft from the car-sharing equation by automating their two basic functions—matching cars with riders and facilitating payments. You could also own "shares" of a car represented by cryptocurrency tokens.
- **PORTABLE MEDICAL DATA:** The same technology that allows two people to exchange units of Bitcoin without necessarily trusting each other could also vouchsafe medical information, putting control firmly in the hands of patients, says Brian Behlendorf, executive director of the Linux Foundation's Hyperledger project, a tool kit for building blockchain applications. Patients would receive a "'health wallet" with their data and histories. A doctor could go to a ledger and request your blood type, generating an access request on the user's phone. "You get an audit trail of who you shared that data with and the option to delete it when the treatment is over," Behlendorf says.
- **A GLOBAL SUPERCOMPUTER:** Linking your devices to thousands of others in a P2P network—and using a blockchain to pay you for their use—would create a financial incentive to support a worldwide, decentralized supercomputer. While you sleep, your laptop and phone could be rented by scientists who want to run models, for example. A project called Golem is already working on it. "The number of idle laptops is so much larger than the computing power of the data centers," Gupta says. "Artificial intelligence, climate modeling—all of that stuff could be accelerated 1,000-fold."

© 2018 Scientific American

WHAT ARE THE LIMITATIONS AND DANGERS OF BLOCKCHAIN?

"Blockchains provide a substrate that, if certain assumptions are held to, is very difficult to modify ex post facto," says Cornell University blockchain researcher Emin Gün Sirer. "But that doesn't mean that everything recorded to a blockchain is true or desirable. If I get hacked and someone steals my cryptocoins and tries to use them, I would very much like to undo that transaction. That's where immutability becomes a liability." It's also easy to confuse a blockchain's theoretical immutability with actual data security: public blockchains like Ethereum and Bitcoin don't actually encrypt any information. The Linux Foundation's Brian Behlendorf goes one step further: "The ledger should never be used to store personal data or anything sensitive, not even in encrypted form," he says, "because we know that no matter what we encrypt today, probably in 40 or 50 years we'll be able to decrypt it" with more advanced technology. Some advocates speak of blockchain as a panacea for any social problem involving trust, but that's blindly optimistic.

HOW DO YOU REGULATE A DECENTRALIZED SYSTEM?

Given the Wild West reputations of decentralized digital currencies, it's easy to assume that they were created to dismantle or avoid financial regulation. But that's not quite accurate. Bitcoin is full of regulations, after all—they're just defined and enforced by source code (and the collective activity of its P2P network) rather than by governments or financial institutions. "The whole innovation about Bitcoin is in eschewing social governance of record keeping," says Patrick Murck, a lawyer who researches blockchain policy and regulation at Harvard University's Berkman Klein Center for Internet and Society. Ethereum's stated purpose—to support the deployment of autonomous smart contracts—is essentially regulatory. A blockchain is arguably nothing *but* regulation: a mathematically enforced system of rules about what can and cannot be done with records in a database.

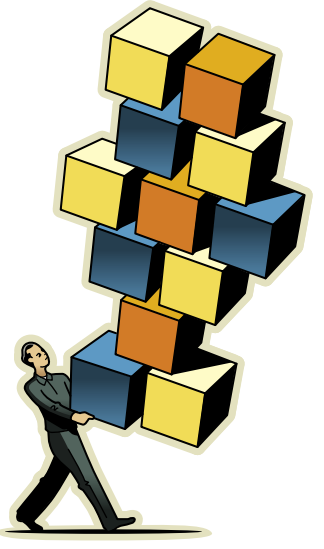
What always matters about financial regulation, decentralized or not, is who gets to do the regulating and how. "If you have a system that's decentralized, there's nowhere to attach regulation—but wherever that system gets reintermediated [by third parties], regulation will follow," Murck says. In 2013 China banned cryptocurrencies from its banking system, and in September 2017 it ordered all domestic Bitcoin exchanges to shut down. The U.S. and Japan are moving to regulate cryptocurrency exchanges and "initial coin offerings" with the same vigilance they apply to stock trading and investment banking.

One future application of blockchain technology is in securing digital identity records, and according to venture capitalist Morris, new cryptocurrencies may emerge that marry identity data with financial information. They wouldn't have the anonymity of Bitcoin (Morris estimates the number of Bitcoin holders who pay honest taxes on them to be mere hundreds). But as digital money goes mainstream, the trade-off in perceived security and stability may make oversight tolerable—or even desirable. Says Murck: "If I'm trusting you with some property to hold on my behalf and transact with it—whether it's Bitcoin or Beanie Babies—then you're either regulated or about to be regulated."

CAN BLOCKCHAINS FAIL?

To date, the Bitcoin blockchain—the world's first and at present its largest and most widely used—has never been compromised or hacked. But that doesn't mean that every blockchain is invulnerable by definition. "There's no such thing as a perfect technology," says Cornell's Gün Sirer, co-director of the Initiative for Cryptocurrencies and Contracts. Here are three gaps in a blockchain's armor:

51% ATTACK: Blockchain-backed cryptocurrency networks rely on two bottomless resources for security: the speed and greed of their miners. But it's theoretically possible to overpower both. To subvert the blockchain's consensus mechanism, hackers would need to gain control over a majority of nodes in the network. This would give them the power to control how and which blocks get mined. They could reverse new transactions, allowing them to double spend digital currency. Or they could prevent other people's transactions from being validated. Bitcoin's P2P network, with thousands of nodes worldwide, seems unlikely to fall prey to such an attack. But smaller "altcoins" are at risk: one called Krypton was hit in 2016 by a group called the 51 crew. Even blockchains that don't use mining are vulnerable because they still rely on an "assumption that a majority of nodes in their network are benign," Gün Sirer warns.



GOOD OLD-FASHIONED HUMAN ERROR: It may take the computing equivalent of moving mountains to compromise the blockchain itself. But anything built on it or attached to it is just as vulnerable as it is now. Mt. Gox, a Bitcoin currency exchange (that is, an intermediary that lets people convert traditional currencies—like dollars—into Bitcoin) plagued by mismanagement and faulty code, lost 850,000 Bitcoins (worth $620 million at the time) in 2014. Ultimately blockchains are just distributed ledgers with no help desk—so if you have a digital wallet full of cryptocurrency and you lose the password, that money is almost certainly gone. There is rich irony in the fact that some cryptocurrency users keep a hard copy of their pass codes (or even the currency itself, stored on a USB drive) in a safety deposit box at the bank—a practice known as cold storage.

"BLOCKCHAIN BLOAT": This is less of a vulnerability than a natural consequence of blockchains working too well. Because every new block essentially revalidates every block before it, that means every node performing the validation needs a copy of the latest version of the entire chain to deal with every new transaction. At more than 130 gigabytes and growing, the Bitcoin blockchain is already getting unwieldy. Ethereum's ledger, designed to be more flexible (so that it can act as a platform for more sophisticated transactions such as smart contracts), is already bigger than Bitcoin's—if everyone were to start using it, would only high-performance supercomputers be able to handle the load? That could effectively decentralize the network, defeating the purpose of the distributed ledger in the first place.

© 2018 Scientific American

© 2018 Scientific American

ARTIFICIAL INTELLIGENCE

CLICKS, LIES AND VIDEOTAPE

Artificial intelligence is making it possible for anyone to manipulate audio and video. The biggest threat is that we stop trusting anything at all

By Brooke Borel

IN BRIEF

Rapidly evolving AI technologies allow for the automated creation of fake video and audio. Some experts worry that the spread of disinformation via social media could have profound effects on public discourse and political stability.

Computer scientists are working on AI detection tools to flag fake videos, but they lag behind the ability to create manipulated content. Meanwhile social scientists warn that policing fakes post hoc is not a sufficient solution.

Written fake news was a troubling factor in the 2016 U.S. elections. Research suggests that fake video may be especially effective at stoking fear—an emotion that powers viral content. One concern is that it could erode our trust in all media, including what is real.

Illustration by Taylor Callery

© 2018 Scientific American

THIS PAST APRIL A NEW VIDEO OF BARACK OBAMA SURFACED ON THE INTERNET. Against a backdrop that included both the American and presidential flags, it looked like many of his previous speeches. Wearing a crisp white shirt and dark suit, Obama faced the camera and punctuated his words with outstretched hands: "President Trump is a total and complete dipshit."

Without cracking a smile, he went on: "Now, you see, I would never say these things. At least not in a public address. But someone else would." The view shifted to a split screen, revealing the actor Jordan Peele. Obama hadn't said anything—it was a real recording of an Obama address blended with Peele's impersonation. Side by side, the message continued as Peele, like a digital ventriloquist, put more words in the former president's mouth.

In this era of fake news, the video was a public service announcement produced by BuzzFeed News, showcasing an application of new artificial-intelligence (AI) technology that could do for audio and video what Photoshop has done for digital images: allow for the manipulation of reality.

The results are still fairly unsophisticated. Listen and watch closely, and Obama's voice is a bit nasally. For brief flashes, his mouth—fused with Peele's—floats off-center. But this rapidly evolving technology, which is intended for Hollywood film editors and video-game makers, has the imaginations of some national security experts and media scholars running dark. The next generation of these tools may make it possible to create convincing fakes from scratch—not by warping existing footage, as in the Obama address, but by orchestrating scenarios that never happened at all.

The consequences for public knowledge and discourse could be profound. Imagine, for instance, the impact on upcoming elections if a fake video smeared a politician during a tight race. Or attacked a CEO the night before a public offering. A group could stage a terrorist attack and fool news outlets into covering it, sparking knee-jerk retribution. Even if a viral video is later proved to be fake, will the public still believe it was true anyway? And perhaps most troubling: What if the very idea of pervasive fakes makes us stop believing much of what we see and hear—including the stuff that is real?

Many technologists acknowledge the potential for sweeping misuse of this technology. But while they fixate on "sexy solutions for detection and disclosure, they spend very little time figuring out whether any of that actually has an effect on people's beliefs on the validity of fake video," says Nate Persily, a law professor at Stanford University. Persily studies, among other topics, how the Internet affects democracy, and he is among a growing group of researchers who argue that curbing viral disinformation cannot be done through technical fixes alone. It will require input from psychologists, social scientists and media experts to help tease out how the technology will land in the real world.

"We've got to do this now," Persily says, "because at the moment the technologists—necessarily—drive the discussion" on what may be possible with AI-generated video. Already our trust in democratic institutions such as government and journalism is ebbing. With social media a dominant distribution channel for information, it is even easier today for fake-news makers to exploit us. And with no cohesive strategy in place to confront an increasingly sophisticated technology, our fragile collective trust is even more at risk.

INNOCUOUS BEGINNINGS

THE PATH TO FAKE VIDEO traces back to the 1960s, when computer-generated imagery was first conceived. In the 1980s these special effects went mainstream, and ever since, movie lovers have watched the technology evolve from science-fiction flicks to Forrest Gump shaking hands with John F. Kennedy in 1994 to the revival of Peter Cushing and Carrie Fisher in *Rogue One*. The goal has always been to "create a digital world where any storytelling could be possible," says Hao Li, an assistant professor of computer science at the University of Southern California and CEO of Pinscreen, an augmented-reality start-up. "How can we create something that appears real, but everything is actually virtual?"

Early on, most graphics came from artists, who used computers to create three-dimensional models and then hand-painted textures and other details—a tedious process that did not scale up. About 20 years ago some computer-vision researchers started thinking of graphics differently: rather than spending time on individual models, why not teach computers to create from data? In 1997 scientists at the Interval Research Corporation in Palo Alto, Calif., developed Video Rewrite, which sliced up existing footage and reconfigured it. The researchers made a clip of JFK saying, "I never met Forrest Gump." Soon after, scientists at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, taught a computer to pull features from a data set of 200 three-dimensional scans of human faces to make a new face.

The biggest recent jump in the relationship among computer vision, data and automation arguably came in 2012, with advanc-

© 2018 Scientific American

TECHNOLOGY that was originally developed to create virtual scenes in film (*1*) has evolved into a tool that can be used to make fake videos (*2*) to spread disinformation.

FILM STILL FROM *FORREST GUMP*. PARAMOUNT PICTURES, 1994 (*1*); FILM STILL FROM *YOU WON'T BELIEVE WHAT OBAMA SAYS IN THIS VIDEO!* MONKEYPAW PRODUCTIONS AND BUZZFEED, APRIL 17, 2018 (*2*)

es in a type of AI called deep learning. Unlike the work from the late 1990s, which used static data and never improved, deep learning adapts and gets better. This technique reduces objects, such as a face, to bits of data, says Xiaochang Li, a postdoctoral fellow at the Max Planck Institute for the History of Science in Berlin. "This is the moment where engineers say: we are no longer going to model things," she says. "We are going to model our ignorance of things and just run the data to understand patterns."

Deep learning uses layers of simple mathematical formulas called neural networks, which get better at a task over time. For example, computer scientists can teach a deep-learning tool to recognize human faces by feeding it hundreds or thousands of photographs and essentially saying, each time, *this is a face* or *this is not a face*. Eventually when the tool encounters a new person, it will recognize patterns that make up human features and say, statistically speaking, *this is also a face*.

Next came the ability to concoct faces that looked like real people, using deep-learning tools known as generative networks. The same logic applies: computer scientists train the networks on hundreds or thousands of images. But this time the network follows the patterns it gleaned from the examples to make a new face. Some companies are now using the same approach with audio. Earlier this year Google unveiled Duplex, an AI assistant based on software called WaveNet, which can make phone calls and sounds like a real person—complete with verbal tics such as uhs and hmms. In the future, a fake video of a politician may not need to rely on impersonations from actors like Peele. In April 2017 Lyrebird, a Canadian start-up, released sample audio that sounded creepily like Obama, Trump and Hillary Clinton.

But generative networks need big data sets for training, and that can require significant human labor. The next step in improving virtual content was to teach the AI to train itself. In 2014 researchers at the University of Montreal did this with a generative adversarial network, or GAN, which puts two neural networks in conversation. The first is a generator, which makes fake images, and the second is a discriminator, which learns to distinguish between real and fake. With little to no human supervision, the networks train one another through competition—the discriminator nudges the generator to make increasingly realistic fakes, while the generator keeps trying to trick the discriminator. GANs can craft all sorts of stuff. At the University of California, Berkeley, scientists built one that can turn images of horses into zebras or transform Impressionist paintings by the likes of Monet into crisp, photorealistic scenes.

Then, this past May, researchers at the Max Planck Institute for Informatics in Saarbrücken, Germany, and their colleagues revealed "deep video," which uses a type of GAN. It allows an actor to control the mouth, eyes and facial movements of someone else in prerecorded footage. Deep video currently works only in a portrait setup, where a person looks directly at the camera. If the actor moves too much, the resulting video has noticeable digital artifacts such as blurred pixels around the face.

GANs are not yet capable of building complex scenes in video that are indistinguishable from ones captured in real footage. Sometimes GANs produce oddities, such as a person with an eyeball growing out of his or her forehead. In February, however, researchers at the company NVIDIA figured out a way to get GANs to make incredibly high-resolution faces by starting the training on relatively small photographs and then building up the resolution step by step. And Hao Li's team at the University of Southern California has used GANs to make realistic skin, teeth and mouths, all of which are notoriously difficult to digitally reconstruct.

None of these technologies are easy for nonexperts to use well. But BuzzFeed's experiment hints at our possible future.

© 2018 Scientific American

The video came from free software called FakeApp—which used deep learning, though not GAN. The resulting videos are dubbed deepfakes, a mash-up of "deep learning" and "fake," named after a user on the Web site Reddit, who, along with others, was an early adopter and used the tech to swap celebrities' faces into porn. Since then, amateurs across the Web have used FakeApp to make countless videos—most of them relatively harmless pranks, such as adding actor Nicolas Cage to a bunch of movies he was not in or morphing Trump's face onto the body of German chancellor Angela Merkel. More ominous are the implications. Now that the technology is democratized, anyone with a computer can hypothetically use it.

CONDITIONS FOR FAKE NEWS

EXPERTS HAVE LONG WORRIED that computer-enabled editing would ruin reality. Back in 2000, an article in *MIT Technology Review* about products such as Video Rewrite warned that "seeing is no longer believing" and that an image "on the evening news could well be a fake—a fabrication of fast new video-manipulation technology." Eighteen years later fake videos don't seem to be flooding news shows. For one thing, it is still hard to produce a really good one. It took 56 hours for BuzzFeed to make the Obama clip with help from a professional video editor.

The way we consume information, however, has changed. Today only about half of American adults watch the news on television, whereas two thirds get at least some news via social media, according to the Pew Research Center. The Internet has allowed for a proliferation of media outlets that cater to niche audiences—including hyperpartisan Web sites that intentionally stoke anger, unimpeded by traditional journalistic standards. The Internet rewards viral content that we are able to share faster than ever before, Persily says. And the glitches in fake video are less discernible on a tiny mobile screen than a living-room TV.

"We will not win this game. It's just that we will make it harder and harder for the bad guys to play it."
—Alexei Efros *University of California, Berkeley*

The question now is what will happen if a deepfake with significant social or political implications goes viral. With such a new, barely studied frontier, the short answer is that we do not know, says Julie Carpenter, a research fellow with the Ethics + Emerging Sciences Group, based at California State Polytechnic University, San Luis Obispo, who studies human-robot interaction.

We have already witnessed the fallout when connectivity and disinformation collide. Fake news—fabricated text stories designed to look like legitimate news reports and to go viral—was a much discussed feature of the 2016 U.S. presidential election. According to collaborative research from Princeton University, Dartmouth College and the University of Exeter in England, roughly one in four Americans visited a fake news site during the five weeks between October 7 and November 14, 2016, mostly through the conduit of their Facebook feeds. Moreover, 2016 marked a low point in the public's trust in journalism. By one estimate, just 51 percent of Democrats and 14 percent of Republicans said they trusted mass media.

The science on written fake news is limited. But some research suggests that seeing false information just once is sufficient to make it seem plausible later on, says Gordon Pennycook, an assistant professor of organizational behavior at the University of Regina in Saskatchewan. It is not clear why, but it may be thanks to "fluency," he says, or "the ease at which it is processed." If we hear Obama call Trump a curse word and then later encounter another false instance where Obama calls Trump obscene names, we may be primed to think it is real because it is familiar.

According to a study from the Massachusetts Institute of Technology that tracked 126,000 stories on Twitter between 2006 and 2017, we are also more likely to share fake news than real news—and especially fake political stories, which spread further and quicker than those about money, natural disasters or terrorism. The paper suggested that people crave novelty. Fake news in general plays to our emotions and personal identity, enticing us to react before we have had a chance to process the information and decide if it is worth spreading. The more that content surprises, scares or enrages us, the more we seem to share it.

There are troubling clues that video may be especially effective at stoking fear. "When you process information visually, you believe that this thing is closer to you in terms of space, time or social group," says Elinor Amit, an assistant professor of cognitive, linguistic and psychological sciences at Brown University, whose work teases out the differences in how we relate to text and images. She hypothesizes that this distinction is evolutionary—our visual development came before written language, and we rely more on our senses to detect immediate danger.

Fake video has, in fact, already struck political campaigns. In July, Allie Beth Stuckey, a TV host at Conservative Review, posted on Facebook an interview with Alexandria Ocasio-Cortez, a Democratic congressional nominee from New York City. The video was not a deepfake but an old-fashioned splice of a real interview with new questions to make Ocasio-Cortez appear to flub her answers. Depending on your political persuasion, the video was either a smear job or, as Stuckey later called it in her defense, satire. Either way, it had 3.4 million views within a week and more than 5,000 comments. Some viewers seemed to think Ocasio-Cortez had bombed a real interview. "Omg! She doesn't know what and how to answer," one wrote. "She is stupid."

That all of this is worrying is part of the problem. Our dark ruminations may actually be worse for society than the videos themselves. Politicians could sow doubt when their real misdeeds are caught on tape by claiming they were faked, for example. Knowing that convincing fakes are even possible might erode our trust in all media, says Raymond J. Pingree, an associate professor in mass communications at Louisiana State University. Pingree studies how confident people are in their ability to evaluate what is real and what is not and how that affects their willingness to participate in the political process. When individuals lose that confidence, they are more likely to fall for liars and crooks, he says, and "it can make people stop wanting to seek the truth."

© 2018 Scientific American

A GAME OF CAT AND MOUSE

TO A COMPUTER SCIENTIST, the solution to a bug is often just more computer science. Although the bugs in question here are far more complex than bad coding, there is a sense in the community that algorithms could be built to flag the fakes.

"There is certainly technical progress that can be made against the problem," says R. David Edelman of M.I.T.'s Internet Policy Research Initiative. Edelman, who served as a tech adviser under Obama, has been impressed by faked videos of the former president. "I know the guy. I wrote speeches for him. I couldn't tell the difference between the real and fake video," he says. But while he could be fooled, Edelman says, an algorithm might pick up on the "telltale tics and digital signatures" that are invisible to the human eye.

So far the fixes fall within two categories. One proves that a video is real by embedding digital signatures, analogous to the intricate seals, holograms and other features that currency printers use to thwart counterfeiters. Every digital camera would have a unique signature, which, theoretically, would be tough to copy.

The second strategy is to automatically flag fake videos with detectors. Arguably the most significant push for such a detector is a program from the Defense Advanced Research Projects Agency called Media Forensics, or MediFor. It kicked off in 2015, not long after a Russian news channel aired fake satellite images of a Ukrainian fighter jet shooting at Malaysia Airlines Flight 17. Later, a team of international investigators pegged the flight's downing on a Russian missile. The satellite images were not made with deep learning, but DARPA saw the coming revolution and wanted to find a way to fight it, says David Doermann, MediFor's former program manager.

MediFor is taking three broad approaches, which can be automated with deep learning. The first examines a video's digital fingerprint for anomalies. The second ensures a video follows the laws of physics, such as sunlight falling the way it would in the real world. And the third checks for external data, such as the weather on the day it was allegedly filmed. DARPA plans to unify these detectors into a single tool, which will give a point score on the likelihood that a video is fake.

These strategies could cut down on the volume of fakes, but it will still be a game of cat and mouse, with forgers imitating digital watermarks or building deep-learning tools to trick the detectors. "We will not win this game," says Alexei Efros, a professor of computer science and electrical engineering at U.C. Berkeley, who is collaborating with MediFor. "It's just that we will make it harder and harder for the bad guys to play it."

And anyway, these tools are still decades away, says Hany Farid, a professor of computer science at Dartmouth College. As fake video continues to improve, the only existing technical solution is to rely on digital forensics experts like Farid. "There's just literally a handful of people in the world you can talk to about this," he says. "I'm one of them. I don't scale to the Internet."

SAVING REALITY

EVEN IF EACH OF US can ultimately use detectors to parse the Internet, there will always be a lag between lies and truth. That is one reason why halting the spread of fake video is a challenge for the social media industry. "This is as much a distribution problem as it is a creation problem," Edelman says. "If a deepfake falls in the forest, no one hears it unless Twitter and Facebook amplify it."

When it comes to curbing viral disinformation, it is not clear what the legal obligations are for social media companies or whether the industry can be regulated without trampling free speech. Facebook CEO Mark Zuckerberg finally admitted that his platform has played a role in spreading fake news—although it took more than 10 months following the 2016 election. Facebook, after all, was designed to keep users consuming and spreading content, prioritizing what is popular over what is true. With more than two billion active monthly users, it is a tinderbox for anyone who wants to spark an enraging fake story.

Since then, Zuckerberg has promised to act. He is putting some of the burden on users by asking them to rank the trustworthiness of news sources (a move that some see as shirking responsibility) and plans to use AI to flag disinformation. The company has been tight-lipped on the details. Some computer scientists are skeptical about the AI angle, including Farid, who says the promises are "spectacularly naïve." Few independent scientists have been able to analyze how fake news spreads on Facebook because much of the relevant data has been on lockdown.

Still, all the algorithms and data in the world will not save us from disinformation campaigns if the researchers building fake-video technology do not grapple with how their products will be used and abused after they leave the lab. "This is my plea," Persily says, "that the hard scientists who do this work have to be paired up with the psychologists and the political scientists and the communication specialists—who have been working on these issues for a while." That kind of collaboration has been rare.

In March, however, the Finnish Center for Artificial Intelligence announced a program that will invite psychologists, philosophers, ethicists and others to help AI researchers to grasp the broader social implications of their work. And in April, Persily, along with Gary King, a political scientist at Harvard University, launched the Social Data Initiative. The project will, for the first time, allow social scientists to access Facebook data to study the spread of disinformation.

With a responsibility vacuum at the top, the onus of rooting out fake videos is falling on journalists and citizen sleuths. Near the end of the deepfake video of Obama and Peele, both men say: "Moving forward, we need to be more vigilant with what we trust from the Internet. It's a time when we need to rely on trusted news sources." It may have been a fake, but it was true. SA

Brooke Borel is a journalist and author of *The Chicago Guide to Fact-Checking* (2016). She recently competed against an AI fact-checker and won by a worrying margin.

MORE TO EXPLORE

The Science of Fake News. David M. J. Lazer et al. in *Science*, Vol. 359, pages 1094–1096; March 9, 2018.

Why Do People Share Fake News? A Sociotechnical Model of Media Effects. Alice E. Marwick in *Georgetown Law Technology Review*, Vol. 2, No. 2, pages 474–512; 2018.

FROM OUR ARCHIVES

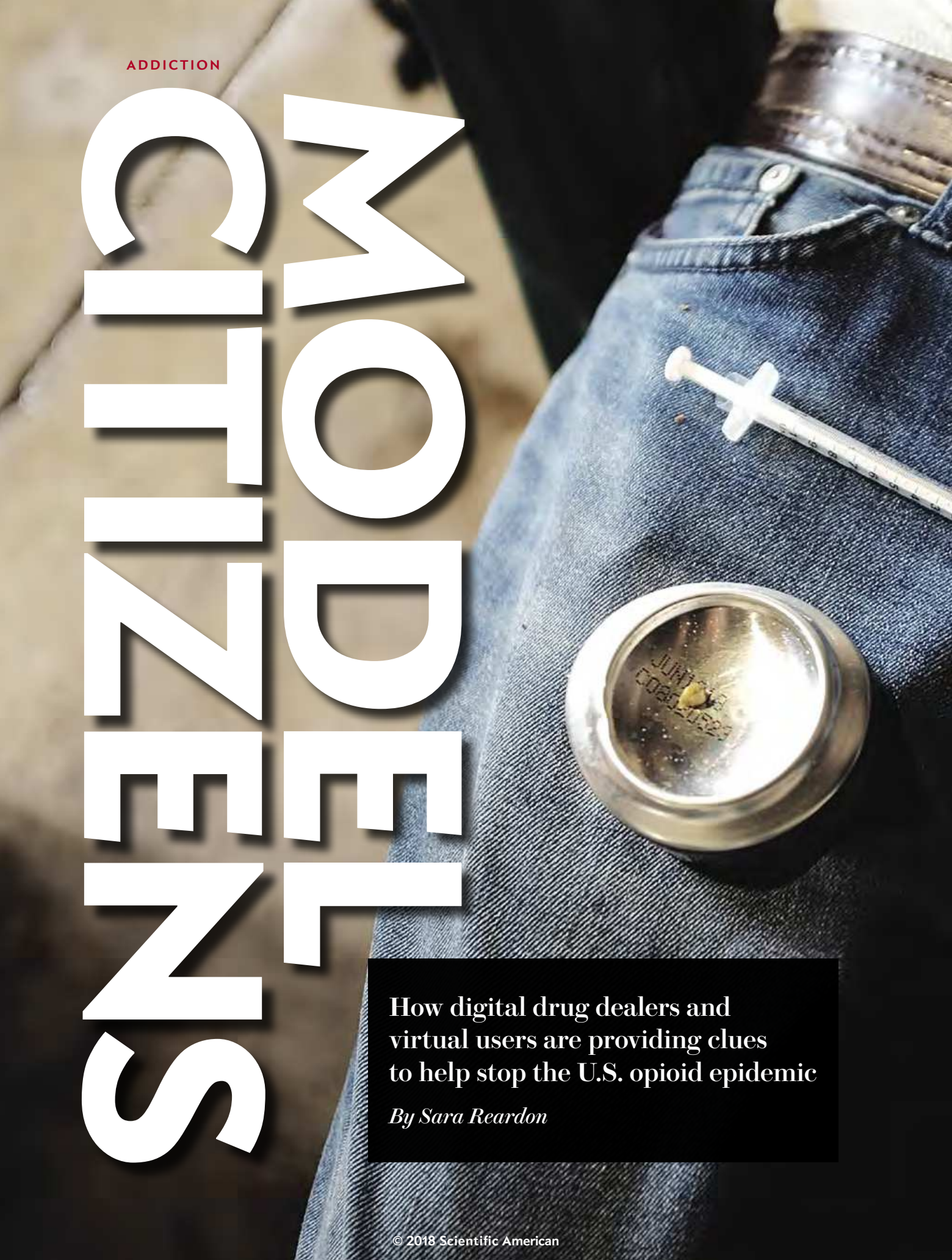
Don't Believe Your Eyes. Lawrence Greenemeier; Advances, April 2018.

scientificamerican.com/magazine/sa

© 2018 Scientific American

ADDICTION

MODELS



How digital drug dealers and virtual users are providing clues to help stop the U.S. opioid epidemic

By Sara Reardon

© 2018 Scientific American

MAN uses heroin under a bridge in Philadelphia.

© 2018 Scientific American

WITH THE TIP OF HER SYRINGE, BRANDI POKES AT A GRAY LUMP of heroin in a spoon. It's a new variety of the drug that has shown up on the market in the past few days, and Brandi likes it. "I feel this more, I feel more of the pain resistance," she says. Once it has dissolved into a liquid, she injects it into her arm, then uses a fresh needle to inject the skinny arm of another woman. "She does it better than the hospital," the woman comments. "I'll help anybody who needs it," Brandi explains to public health researcher Daniel Ciccarone of the University of California, San Francisco, who has been filming the entire process.

Ciccarone's team has embedded with Brandi—whose name has been changed for this story—in Charleston, W.Va., documenting her interactions without judgment or interference. Later, the group will analyze this video, in addition to half a dozen other videos of drug users from across the city, logging details big and small. Brandi does not heat the solution on the spoon, for instance, and that may increase the likelihood of spreading viruses such as HIV. And tests reveal that what she's taking has been laced with fentanyl, a synthetic drug up to 50 times more powerful than heroin.

The researchers will plug these data into powerful computer simulations of Charleston, populated by thousands of virtual Brandis—heroin users and dealers going about their daily routines. They will watch these digital agents buy more heroin as their tolerance increases, form networks with sellers and users, and, in some cases, accidentally overdose.

Ciccarone's is one of several groups using agent-based models to understand what is driving the U.S. opioid epidemic—the dramatic rise over the past two decades in the use of opioids, including prescription pain medications and illegal drugs such as heroin. By studying the motivations and practices of real drug dealers and users, the researchers hope to build agents whose behavior in the virtual world mimics that in real life.

Agent-based models promise to provide a more granular view of the opioid crisis than standard modeling, which is based on average populations, and to capture some of the complexity of the driving forces. This could prove important for demonstrating the effects of opening or closing methadone clinics or needle exchanges. The models allow scientists to compare interventions at almost no cost and could help policy makers to decide how to proceed in the real world. "It's a very classic and useful way to try and see where is the best place to deploy an intervention to have the biggest effect," says John Brooks, a medical adviser for the division of HIV/AIDS prevention at the U.S. Centers for Disease Control and Prevention.

Although such simulations have long been used to model disease outbreaks and have, in some instances, guided public policy, their track record with more complex social behavior such as drug use is limited, largely because of sparse data and the breadth of parameters to consider.

Still, scientists hope that agent-based models can lay out scenarios for decision makers, who are often driven more by politics than data. "The barriers are not scientific or medical," Ciccarone says. "You can throw $1 billion at West Virginia, and they may or may not know how to use it well." These virtual worlds can add clarity, says Joshua Epstein, director of the Agent-Based Modeling Lab at New York University. "You can literally watch the thing unfold before your eyes," he says.

THE DIFFERENCE IN THE DETAILS

THE U.S. OPIOID CRISIS is estimated to kill 115 people a day through overdoses and has run up $1 trillion in health care costs and lost productivity since 2001. It is not the first addiction crisis that the U.S. has faced, nor is it the most severe. Alcohol use causes many more deaths, and the rate of cocaine overdose among African-Americans is similar to the rate of opioid overdose in white Americans.

PRECEDING PAGES AND OPPOSITE PAGE: SPENCER PLATT *Getty Images*

© 2018 Scientific American

WARREN, OHIO, has high rates of unemployment and opioid addiction. Here an American flag is painted inside the window of a closed business.

But the opioid crisis does have some different driving factors, including the prevalence of prescription drugs, which many have used on the way to abusing illegal drugs, and the introduction of fentanyl, which is often used to boost the potency of heroin and is responsible for a large share of overdose deaths. The epidemic has also hit hard in rural settings, where services and infrastructure for dealing with addiction are scarce. "The demographic now encompasses a population that in the past has not been so affected," says Nora D. Volkow, director of the National Institute on Drug Abuse.

As a result, researchers are coming up with fresh ways of thinking about the crisis. It bears similarities to a disease epidemic, for example, in the way it spreads through networks based on personal relationships and physical proximity, says Georgiy Bobashev, a data scientist at the nonprofit research institute RTI International in Research Triangle Park, N.C. "Nobody is born an addict. Somebody has to teach you how to smoke or how to inject."

These personal networks can be replicated using agent-based modeling. Unlike other types of model, which may rely on average characteristics or relationships between homogeneous groups to inform algorithms, agent-based models allow researchers to see subtle connections between people. "That's useful because drug use and overdose is inherently personal," says epidemiologist Brandon Marshall of Brown University. Factors such as job loss, mental health or genetics can influence how likely a person is to begin using drugs or become addicted, but those factors might fade into the averages if researchers looked at a population as a whole.

To create an agent-based model, researchers first "build" a virtual town or region, sometimes based on a real place, including buildings such as schools and food shops. They then populate it with agents, using census data to give each one its own characteristics, such as age, race and income, and to distribute the agents throughout the virtual town.

The agents are autonomous but operate within preprogrammed routines—going to work five times a week, for instance. Some behaviors may be more random, such as a 5 percent chance per day of skipping work or a 50 percent chance of meeting a certain person in the agent's network. Once the system is as realistic as possible, the researchers introduce a variable such as a flu virus, with a rate and pattern of spread based on its real-life characteristics. They then run the simulation to test how the agents' behavior shifts when a school is closed or a vaccination campaign is started, repeating it thousands of times to determine the likelihood of different outcomes.

In 2015 data from an agent-based model developed at the University of Pittsburgh helped California state senator Richard Pan to gain support for a bill on mandatory vaccination in his state. Pan used the simulation to demonstrate to his fellow senators how measles outbreaks could unfold in their home districts. "It certainly made an impact on them," Pan says. "Instead of just describing it in more abstract terms, [the model] can

© 2018 Scientific American

LAW-ENFORCEMENT OFFICIALS seize 33 kilograms of heroin and two kilograms of fentanyl in a major drug bust in New York State.

DREW ANGERER Getty Images

make it more concrete." The bill ultimately passed, and immunization rates increased.

As computers have improved, researchers have begun adapting agent-based models to look at sociological and behavioral trends that require more computing power because of the number of variables they contain. Some groups use the technique for crisis modeling, and Australia has begun intervention studies for child obesity on the basis of the findings of an agent-based model.

In response to the opioid epidemic, Bobashev's group has constructed Pain Town—a generic city complete with 10,000 people suffering from chronic pain, 70 drug dealers, 30 doctors, 10 emergency rooms and 10 pharmacies. The researchers run the model over five simulated years, recording how the situation changes each virtual day.

During this time, the patients' drug tolerance increases, leading them to find different ways of acquiring drugs. Their behavior is driven by variables such as the chance that a doctor will increase their prescription or the likelihood that a dealer will have enough heroin. At a certain threshold, patients become addicted or more likely to overdose. Bobashev's early data suggest, for example, that requiring doctors to track patients' medication history can be effective over the long term, though not immediately.

The model contains many assumptions and simplifications, Bobashev says. For example, it does not capture the fact that the rate at which people develop tolerance and addiction can depend on factors such as genetics and that whether a person switches from prescription drugs to heroin can depend on the relative availability of the two drugs.

But researchers can adjust models such as Pain Town to test various interventions, such as increasing access to emergency rooms, arresting a dealer or equipping police with naloxone (a drug that reverses opioid overdoses), to see how the system reacts and whether it affects the number of deaths over time. And as models become more sophisticated, the researchers may be able to incorporate more factors, such as people who are not taking pain medications but are susceptible to trying opioids for the first time.

Models can also be useful for understanding why individual places or situations may differ, says Christopher Barrett, a computer scientist at Virginia Tech. For instance, heroin and fentanyl might be easier to come by in cities near ports, whereas doctors may be the main source of opioids in a suburban or rural setting. Interventions focused on prescribing practices, therefore, would have different effects in each case.

Such models can also reveal feedback loops, such as the link between economic downturns and opioid use. Some epidemiological studies have suggested that factors such as unemployment tend to predict suicide and addiction, especially in white

© 2018 Scientific American

male populations. And addiction can lead to further job loss and lower productivity, harming the economy. Agent-based models could investigate loops such as this, providing ideas for how to mitigate the effects, Barrett says.

In May, Bobashev and Ciccarone presented results from one of their agent-based models at a meeting of the International Society for the Study of Drug Policy in Vancouver, B.C. Their findings suggested that the increased prevalence of white-powder heroin—a newer form of the drug in the U.S.—may increase the risk of HIV spreading among injection drug users. The reason, also supported by the model, is that unlike black-tar heroin, users do not need to heat the drug to dissolve it—and heating kills the virus.

Bobashev and Ciccarone are working on models of how younger heroin users begin using the drug. Unlike older users, who experienced the rise of the HIV epidemic in the 1980s, newer users may be less likely to adopt safe practices. The models suggest that the U.S. may see more localized HIV outbreaks, similar to the recent outbreak in Scott County, Indiana. That region experienced 181 new HIV cases between November 2014 and November 2015, compared with fewer than five cases per year previously. Opioid use is thought to be the cause. Agent-based models might help stem future outbreaks by guiding surveillance priorities.

Law-enforcement officials have been seizing vast quantities of heroin and the powerful synthetic drug fentanyl, but information on the amount in circulation is hard to come by.

One of the most sophisticated agent-based models is the University of Pittsburgh's system, known as FRED (a Framework for Reconstructing Epidemiological Dynamics). It fits population census data to maps of geographical regions around the country, allowing researchers to track virtual individuals in the area in a realistic way. It was data from these models that helped to convince Pan and his fellow state senators to pass legislation on mandatory vaccination. The FRED team is now beginning to use the system for opioid modeling, training it on historical trends. Pan, who is also a physician, says he is intrigued by the prospect. "If there's a way to actually model in different communities which factors would have the biggest impact, that would be helpful," he says.

DATA DROUGHT

THE MODELS FACE numerous challenges before they will be ready for widespread adoption, primarily data gaps. Marshall says that researchers struggle to get access to data on opioid prescriptions that are held by manufacturers, pharmacies and law-enforcement agencies. It is also difficult to obtain government information on drug cartels and the type and rate of drugs flowing into the country. Other data simply do not exist in usable form: agencies may record deaths from drug overdose, for instance, but fail to specify which drug was responsible.

Observing drug users such as Brandi can provide certain types of information more quickly and accurately. "Drug users know their chemicals intimately," Ciccarone says.

Lee Hoffer, a cultural anthropologist at Case Western Reserve University, studies heroin markets and collaborates with Bobashev. He says that the ethnographic data that his group and others are collecting could help fill some of the information gaps: "We're trying to enter their world as interlopers to see how they see their life." After an initial awkward period, he notes, drug users tend to become more honest with the researchers, telling them crucial information such as how they form networks with dealers and the cost of drugs.

Understanding the psychology of drug users is also crucial, Epstein says. Most decision-making models assume rational behaviors. In reality, emotions, misinformation and irrational calculations play a major part. "When you put them together, you get collections of dynamics that are very dysfunctional."

Epidemiological data may soon be available to buttress the models. The CDC and the National Institute on Drug Abuse have started several major surveys of drug-use patterns. A number of states have also begun collecting epidemiological information on trends in overdose and addiction. And research groups such as the University of Pittsburgh team are working with multiple health agencies to collate their findings in a single database, which can inform FRED and other models.

But no matter how advanced the models become, implementing interventions based on their findings is an enormous challenge. Models may reveal socioeconomic contributors that cannot be easily addressed by policies, and politics can stand in the way of proven solutions. In April, Ciccarone had to cancel his work in Charleston, at least for the time being, after a needle-exchange clinic with which he had been collaborating closed because of political pressure. "They were seeing 300 people on a Wednesday afternoon because there's a lot of need," he says. "It's a huge loss."

Increasing work is being done to determine the relative impact of interventions. In April the National Institutes of Health announced $96 million for a program that will partner with health care systems and local governments to carry out evidence-based public health interventions in different locations, evaluating them as they go. "This is the first time this [has been] done for a particular substance-abuse disorder," Volkow says. The NIH is now asking researchers who want to apply for these funds to justify the size and scope of their proposed studies with data from models, including agent-based models.

But these studies are certain to take many years to complete. And Bobashev says that society cannot afford to wait for the science to be perfect: "By the time this data is collected, tens of thousands, if not hundreds of thousands, more deaths will have occurred." SA

Sara Reardon is a senior reporter at *Nature* in Washington, D.C., covering biomedical research and policy. She has previously written for *New Scientist* and *Science* and has a master's degree in molecular biology from the University of Washington.

MORE TO EXPLORE

FRED (a Framework for Reconstructing Epidemic Dynamics): An Open-Source Software System for Modeling Infectious Diseases and Control Strategies Using Census-Based Populations. John J. Grefenstette et al. in *BMC Public Health,* Vol. 13, No. 940; October 8, 2013.

Pain Town, an Agent-Based Model of Opioid Use Trajectories in a Small Community. Georgiy Bobashev et al. in *Social, Cultural, and Behavioral Modeling.* Edited by R. Thomson et al. Lecture Notes in Computer Science, Vol. 10,899. Springer, 2018.

FROM OUR ARCHIVES

A Lifeline for Addicts. Michele Solis; *Scientific American Mind,* March/April 2013.

scientificamerican.com/magazine/sa

© 2018 Scientific American

END NOTE

A Good Life for All

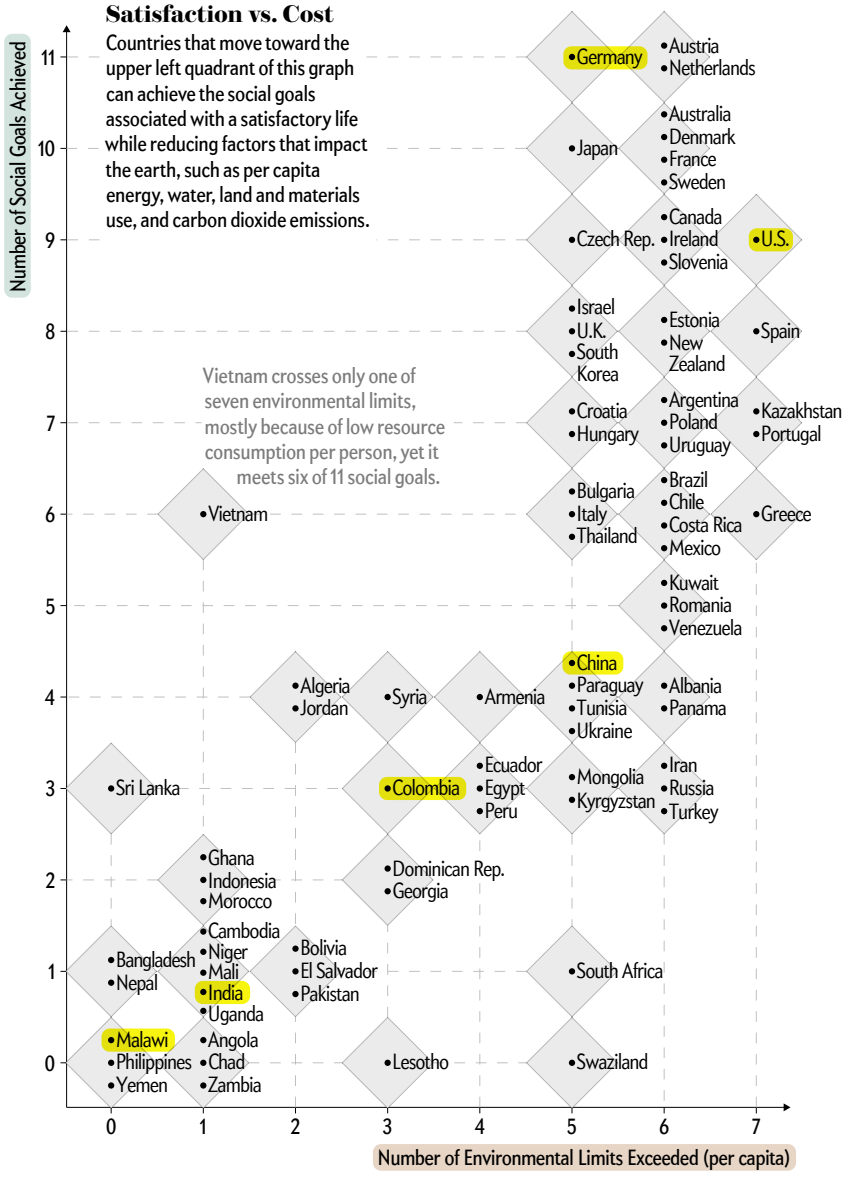
Can humans live well without pillaging the planet?

Text by Mark Fischetti
Graphic by Federica Fragapane

Many wealthy nations achieve a range of social objectives that together can provide a good life for their people, as outlined by the United Nations Sustainable Development Goals. But to do so, they exceed their share of the earth's natural resources and surpass environmental impact limits needed to safeguard the planet, according to a recent study (*top right of main graph*). Less wealthy nations use resources more modestly and have lower impacts but meet fewer of the social goals (*bottom left of main graph*). The solution: "Wealthy nations can consume less, with no loss in quality of life," says study leader Daniel W. O'Neill of the University of Leeds in England. That would free up resources for less wealthy nations to improve lives (*circular charts*) while still keeping within safe environmental boundaries.

Satisfaction vs. Cost

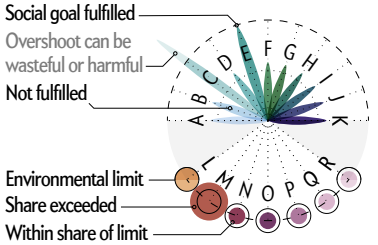
Countries that move toward the upper left quadrant of this graph can achieve the social goals associated with a satisfactory life while reducing factors that impact the earth, such as per capita energy, water, land and materials use, and carbon dioxide emissions.



Country Goals and Limits

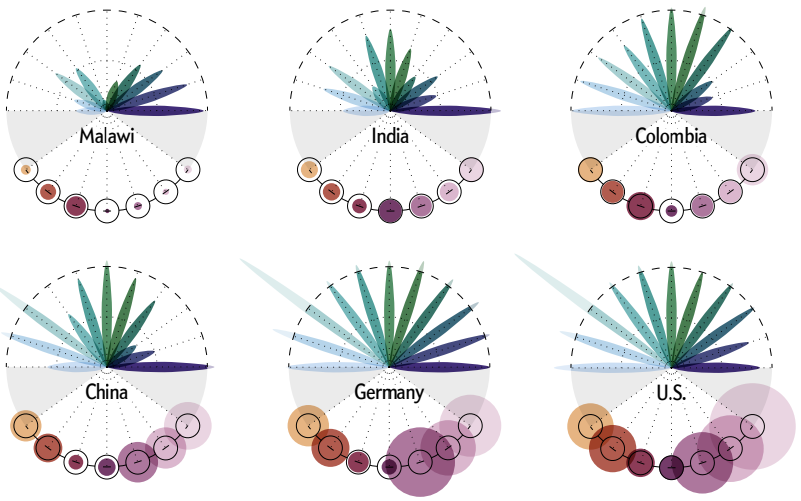
Social Goals

- A Life satisfaction
- B Years of healthy life
- C Nutrition
- D Sanitation
- E Sufficient income
- F Access to energy
- G Education
- H Social support
- I Democratic rights
- J Income equality
- K Employment rate


Social goal fulfilled
Overshoot can be wasteful or harmful
Not fulfilled
Environmental limit
Share exceeded
Within share of limit
A B C D E F G H I J K
L M N O P Q R

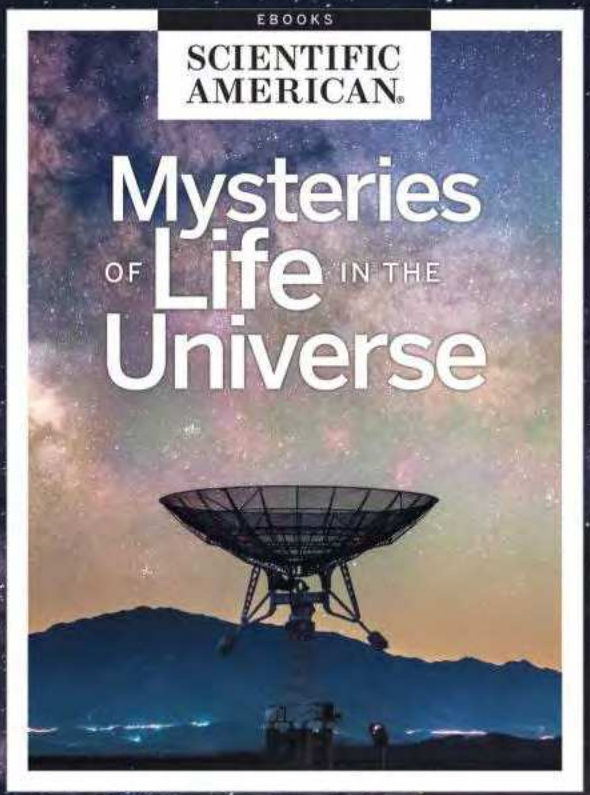
Environmental Limits

- L Materials use
- M Land and ocean use
- N Crop and forest loss
- O Freshwater use
- P Nitrogen discharge
- Q Phosphorus discharge
- R CO_2 emissions



SOURCE: "A GOOD LIFE FOR ALL WITHIN PLANETARY BOUNDARIES," BY DANIEL W. O'NEILL ET AL., IN *NATURE SUSTAINABILITY*, VOL. 1; FEBRUARY 5, 2018 *https://goodlife.leeds.ac.uk*

© 2018 Scientific American



Examine the Big Questions

How did life begin?
Are we alone in the universe?
In this eBook, we explore several theories on the origin of life, some of its extreme forms and the ongoing search for signs – or sentience – on distant worlds.

sciam.com/ebooks

SCIENTIFIC AMERICAN eBooks

Scientific American is a registered trademark of Springer Nature America, Inc.

SCIENTIFIC AMERICAN | Travel

Great Britain

August 13-25, 2019

Join fellow science buffs for a unique vacation that will inspire, educate and entertain with:

Seminars at Sea. Learn cutting-edge science between ports of call

Expert Speakers. Discuss the latest breakthroughs with pioneers in their fields

Social Events. Enjoy the intimate community feel at group receptions, dinners and more

Sightseeing. Go museum-hopping, hiking, biking, shopping or take a walking tour

Broaden your horizons.

For more information email us at **info@InsightCruises.com** or visit **ScientificAmerican.com/travel**

Cruise prices start at $2,809 per person (pp) based on double occupancy. For those attending our SEMINARS, there is a $1,575 fee. Add'l pp fees: gov't taxes and fees ($210), a non-refundable Booking Service Fee ($150), and the Tour Leader gratuities ($10 pp per day). Program, cruise pricing and options are subject to change.